Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/010733

International filing date: 29 March 2005 (29.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/557,929
Filing date: 30 March 2004 (30.03.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1352370

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/557,929*
FILING DATE: *March 30, 2004*
RELATED PCT APPLICATION NUMBER: *PCT/US05/10733*

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

ATTENTION: PROVISIONAL PATENT APPLICATION

Sir:

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR § 1.53(c).

For: **EXPRESSION ANALYSIS OF BRONCHIAL CELLS TREATED WITH TOBACCO CONDENSATES**

Name of First Inventor: Ellen D. Jorgensen

Name of Second Inventor: Igor Dozmorov

Name of Third Inventor: Mark B. Frank

Name of Fourth Inventor: Michael Centola

Name of Fifth Inventor: Anthony P. Albino

Enclosed are:

(X) Specification in 65 pages.

(X) A check in the amount of $160 to cover the filing fee is enclosed.

(X) A return prepaid postcard.

(X) The Commissioner is hereby authorized to charge any additional fees which may be required, now or in the future, or credit any overpayment to Account No. 11-1410.

Was this invention made by an agency of the United States Government or under a contract with an agency of the United States Government?

(X) No.

( ) Yes. The name of the U.S. Government agency and the Government contract number are:

(X) Please send correspondence to:

Eric S. Furman, Ph.D.
Knobbe, Martens, Olson & Bear, LLP
2040 Main Street, 14th Floor
Irvine, CA 92614

Respectfully submitted,

Eric S. Furman, Ph.D.
Registration No. 45,664
Customer No. 20,995
(619) 235-8550

S:\DOCS\ESF\ESF-7686.DOC
033004

Knobbe Martens Olson & Bear LLP
Intellectual Property Law

2040 Main Street
Fourteenth Floor
Irvine, CA 92614
Tel 949-760-0404
Fax 949-760-9502
www.kmob.com

Eric S. Furman, Ph.D.

MAIL STOP PROVISIONAL PATENT APPLICATION
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

Attorney Docket No.	:	VTOB.302PR
Applicant(s)	:	Jorgensen et al.
For	:	EXPRESSION ANALYSIS OF BRONCHIAL CELLS TREATED WITH TOBACCO CONDENSATES
Attorney	:	Eric S. Furman, Ph.D.
"Express Mail" Mailing Label No.	:	EV 323822229 US
Date of Deposit	:	March 30, 2004

I hereby certify that the accompanying

Transmittal letter; Specification in 65 pages; Check(s) for Filing Fee(s); Return Prepaid Postcard

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and are addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Printed Name: _____

Knox Services
2250 Fourth Avenue
Suite 200, San Diego, CA 92101

ISMAEL CENTENO

S:\DOCS\ESF\ESF-7687.DOC
033004

Global Gene Expression Analysis of Human Bronchial Epithelial Cells Treated with Tobacco Condensates

Ellen D. Jorgensen[1], Igor Dozmorov[2], Mark B. Frank[2], Michael Centola[2], Anthony P. Albino[1,3].

[1]Vector Research Ltd.

712 Fifth Ave., New York, NY 10019

[2]Oklahoma Medical Research Foundation

Department of Arthritis and Immunology

825 NE 13th St., Oklahoma City, OK 73104

[3]To whom all correspondence should be addressed at:

Vector Research Ltd.

712 Fifth Ave., New York, NY 10019

phone: 212 409-2800

Fax: 212 409-2801

Email: talbino@vectorgroupltd.com

Keywords: cigarette smoke, microarray, S9 toxicity; bronchial cell

Running Title: Expression Analysis of Bronchial Cells Treated with Tobacco Condensates

INTRODUCTION

The leading preventable cause of death and disability in the United States is the chronic use of tobacco products, in particular, cigarettes (40; 91). In addition to lung cancers, tobacco use plays important direct and indirect roles in the etiology of a wide range of other cancers, including those of the upper aerodigestive tract (i.e., oral cavity, pharynx, larynx, and esophagus), kidney, stomach, bladder, pancreas, uterine cervix, and blood (i.e., certain leukemias). Exposure to tobacco carcinogens and toxins is also a major cause of other diseases of the pulmonary system (i.e., bronchitis, emphysema, chronic obstructive pulmonary disease), the cardiovascular system (i.e., stroke, atherosclerosis, and myocardial infarction), and the female reproductive system (i.e., increased risk of miscarriage, premature delivery, low birth weight, stillbirth, and infant death). While numerous studies have elucidated some of the chemical and biological properties of cigarette smoke that result in its ability to induce this range of pathologies in the smoker, little is known about the nature and temporal association of molecular events that drive specific stages in the multi-step processes that result in clinically evident disease (40). This is due, in part, to the limited number of individual tobacco constituents such as benzo[a]pyrene that have been assessed for genetic impact, and the fact that few studies have attempted to address the synergistic relationships between the thousands of individual compounds that constitute the various classes of carcinogens in the vapor and particulate phases of tobacco smoke on gene expression (75). Cigarette smoke is primarily a mixture of gases (i.e., nitrogen, oxygen, and carbon dioxide) and suspended particulate material that consists of a wide variety of condensed organic compounds (i.e., 'tar'). This particulate phase contains the majority of compounds [at least 60] for which there is sufficient evidence of carcinogenic potential in animals or humans (40; 43; 45). Presumably, the inherent chemical complexity of cigarette smoke results in an equally complex biological response involving a

number of signaling pathways and checkpoints that respond to the direct and indirect stress on the genome in exposed tissues. Thus, assessing global gene expression patterns using high-density microarrays is an especially useful approach to detail how a lung cell mounts a multigenic response to cigarette smoke and to the major classes of constituents (e.g., vapor and particulate phases) comprising cigarette smoke. Therefore, the current study determined the impact of different CSCs on global gene expression profiles in short-term cultures of NHBE cells.

Analysis of the data from these types of large-scale gene expression studies is nontrivial due to the complexity and size of data sets and the fact that technical variation can be introduced at different stages in array production and processing. Establishing well-specified and carefully validated procedures for standardization and normalization of array data from individual specimens is a key step in the analysis. However, no current single method has proven free from ambiguity. Selection criteria based on the ratio of measured expression levels fails to account for intra-group variations (i.e. normal biologic variance) and can result in false positive selections (22; 47). Additionally, current statistical methods do not adequately address the mutually exclusive characteristics of sensitivity and specificity. The common practice of using low thresholds for selection of significance ($p<0.05$) can also result in a large number of false positive selections. This is especially problematic for high-density arrays as the number of false positive selections expected to occur by chance may limit the ability to perform higher order analyses, such as molecular pathway identification or disease subphenotyping, that require groups of differentially expressed genes to be accurately predicted. Attempts to increase stringency by raising the threshold of significance above this value can also be problematic, as it will cause a compensatory decrease in sensitivity and resultant increase in false negative selections. The use of large numbers of replicates is able improve this situation (33), although it

can be expensive and labor intensive. We addressed many of these issues by previously developing (23; 33) (and which we describe further in this paper) a novel method of microarray data mining, denoted hypervariable analysis, which uses statistical robust delimiters for defining biologically-relevant changes in gene expression. Hypervariable analysis is predicated on the observation that a biologically relevant stimulus will alter gene expression such that homeostasis of the transcriptome is disrupted. Accordingly, these stimuli will modulate the levels of mRNAs of affected genes such that their expression variance over time exceeds the variance observed in the majority of genes in an unstimulated state. While hypervariability has not been previously defined per se, numerous examples of the importance of this parameter of gene behavior exist. The most compelling examples can be seen in time-course studies of yeast cells in which the mitosis-related genes become demonstrably hypervariable (80). Examining subtle alterations to the 'homeostatic transcriptome' may be useful in defining the major signaling pathways activated upon exposure to chronic, but low level, doses of carcinogenic mixtures such as occurs daily in an individual smoker. This type of analysis is especially relevant for complex bioactive mixtures such as cigarette smoke since assessing the specific effects of individual components of such mixtures may not reflect their true impact due to synergistic or antagonistic interactions with other components normally present. Moreover, cigarette smoke (as opposed to a single agent with a well-defined mechanism of toxicity) might be expected to result in multiple types of genomic and cellular damage over time since both the vapor and particulate components of tobacco smoke contain numerous substances that immediately and directly damage a range of biomolecules, as well as other substances whose toxicity is activated only after biotransformation by cellular enzymes into reactive nucleophiles that then attack various cellular elements.

In this paper, the method of hypervariable analysis is paired with an experimental design, specifically a time-course analysis, to characterize the overall patterns of change in the

'homeostatic transcriptome' of short-term cultures of NHBE cells treated with different CSCs over a period of 12 hours. This analysis showed that 1) exposure of NHBE cells to CSCs from different commercial brands of cigarettes alters the expression of a large common set of genes in both a transitory and sustained manner; 2) each CSC also affects the expression of a smaller non-overlapping set of unique genes; 3) both CSCs impact genes that participate in a diverse set of biological pathways whose dysfunction is relevant to a number of diseases including cancer; and 4) the S9 metabolic fraction of enzymes has a significant impact on gene expression which differs from that of both CSCs. The identification of tobacco-affected gene sets, as well as the biological phenomena in which these genes participate, will advance the generation of a detailed atlas of molecular events caused by exposure to tobacco smoke constituents. This atlas will be invaluable for clarifying the relationship between altered gene expression and cellular dysfunction, an important step in developing a highly accurate model of disease risk for current and former users of tobacco products.

MATERIALS AND METHODS

Preparation of cigarette condensates: Smoke was generated from two commercially available nationally sold brands of American cigarettes (Brand A and Brand B) using an INBIFO-Condor smoking machine under Federal Trade Commission (FTC) smoking parameters (2.0 second puff duration, 35 milliliter puff every 60 seconds) (26). Both brands of cigarettes are non-menthol, full-flavor types of American-blended cigarettes with averaged FTC measured values of 13.2 mg tar/0.88 mg nicotine (Brand A), and 14.5 mg tar/1.04 mg nicotine (Brand B). Smoke condensates extracted from these two cigarette brands and designated CSC-A and CSC-B, respectively, were collected from the smoke via a series of three cold traps (-10°C, -40°C, and -70°C) onto impingers filled with glass beads. The condensates were dissolved in acetone, which was then removed by rotary evaporation at 35°C. The resulting CSCs were weighed and dissolved in dimethylsulfoxide (DMSO) to make stock solutions of each condensate at a concentration of 40 mg/mL, which were stored at -20°C prior to use.

Cell Culture and Treatment: NHBE cells were purchased from Cambrex Corporation, East Rutherford, NJ. The cells were cultured in complete Bronchial Epithelial Cell Growth Medium (BEGM), prepared by supplementing Bronchial Epithelial Basal Medium with retinoic acid, epidermal growth factor, epinephrine, transferrin, T3, insulin, hydrocortisone, antimicrobial agents and bovine pituitary extract by addition of SingleQuots,™ (both purchased from Cambrex Corporation, East Rutherford, NJ). S9 metabolic fraction from Aroclor 1254-treated rats was obtained from BioReliance Corporation, Rockville, MD. A 5x concentration of S9 fraction with cofactors was prepared immediately before treating the cells, and contained 10% S9, 4mM NADP, 5 mM glucose-6-phosphate, 50mM phosphate buffer pH 8.0, 30 mM KCl, and 10 mM $CaCl_2$.

Twenty-eight flasks were seeded with 14.6 ml of a 2.52 x 10^4 cells/ml cell suspension and an additional 15.4 ml pre-warmed BEGM were added to each flask for a final volume of 30 mL/flask. All incubations were at 37°C in a humidified atmosphere of 5% CO_2 in air. Cells were grown to 40% confluence, at which time the cultures were treated. Four flasks were used as untreated control cultures. Following medium removal in these four control flasks, the cells were refed with 30 ml pre-warmed BEGM and their RNA harvested at 0h (2 flasks) and 20 hr (2 flasks). The remaining 24 experimental flasks were treated with either CSC-A in the presence of 2% S9 fraction, CSC-B in the presence of 2% S9 fraction, or 2% S9 fraction alone. Following medium removal, each flask received 9.0 ml of fresh BEGM, 15.0 mL BEGM containing CSC or vehicle (400 μg/ml of CSC-A or CSC-B and 1% DMSO for the CSC-treated groups, 15.0 mL containing 1% DMSO for the S9-only group), and 6 ml of 5x S9 fraction for a final concentration of 2% S9 and a final media volume of 30 mL. Incubation was carried out under the incubation conditions described above. Duplicate flasks were used for each treatment/time point of the experiment (i.e., 2, 4, 8, and 12h).

RNA Preparation. Cells were harvested for total RNA extraction after 0 (untreated), 2, 4, 8, and 12 hours of treatment. The medium was aspirated and the flasks were rinsed twice with prewarmed 15 mL Dulbecco's Phosphate Buffered Saline. After the second rinse, 5.0 mL of cold TRIzol® (Invitrogen Corp., Carlsbad, CA) were added to cover the cells in each flask. Each flask was vigorously vortexed for approximately one minute. The TRIzol® was pipetted up and down over the surface of the flask at least five times to suspend the cell lysate. The resulting TRIzol®/cell lysate was allowed to remain in the flask for at least 10 minutes at room temperature after which it was transferred to microfuge tubes and extracted with 0.2 ml chloroform per 1.0 ml TRIzol/cell lysate. The tubes were capped and shaken vigorously to

initiate the RNA extraction, and centrifuged at >15,000 x g for two 5-minute spins. Following the second 5-minute centrifugation, the aqueous layer was collected (~500 µl) and transferred to a second set of microfuge tubes containing an equal volume of isopropyl alcohol. The samples were centrifuged for 30 minutes at >15,000 x g. Following centrifugation, most (~90%) of the liquid was removed from the microfuge tube. The remaining RNA pellet was frozen and stored at <-60°C. RNA was resuspended in diethylpyrocarbonate-treated water. RNA integrity was assessed using capillary gel electrophoresis (Agilent Technologies, Palo Alto, CA) to determine the ratio of 28s:18s rRNA in each sample.

Microarray Printing and Processing: The microarrays used in these experiments were developed at the Oklahoma Medical Research Foundation Microarray Research Facility. Slides were produced using commercially available libraries of 70 nucleotide long DNA molecules whose length and sequence specificity were optimized to reduce the cross-hybridization problems encountered with cDNA-based microarrays (Human Genome Oligo Set Version 2.0, Qiagen, Valencia, CA). The microarrays had 21,329 human genes represented. The oligonucleotides were derived from the UniGene and RefSeq databases. The RefSeq database is an effort by the NCBI to create a true reference database of genomic information for all genes of known function. For the genes present in this database, information on gene function, chromosomal location, and reference naming are available. All 11,000 human genes of known or suspected function are represented on these arrays. In addition, most undefined open reading frames were represented (approximately 10,000 additional genes). Oligonucleotides were resuspended at 40µM concentrations in 3xSSC and spotted onto Corning® UltraGAPS™ amino-silane coated slides, rehydrated with water vapor, snap dried at 90°C, and then covalently fixed to the surface of the glass using 300 mJ, 254 nm wavelength ultraviolet radiation. Unbound free

amines on the glass surface were blocked for 15 min with moderate agitation in a 143 mM solution of succinic anhydride dissolved in 1-methyl-2-pyrolidinone, 20mM sodium borate, pH 8.0. Slides were rinsed for 2 min in distilled water, immersed for 1 min in 95% ethanol, and dried with a stream of nitrogen gas.

cDNA Synthesis and Hybridization: cDNA was synthesized with a direct incorporation of Cy3-dUTP from 2 ug total RNA using Clontech Powerscript (Clontech, Palo Alto, CA) reverse transcriptase. Labeled cDNA was purified using a Montage 96-well vacuum system. The cDNA was added to hybridization buffer containing Cot-1 DNA (0.5 mg/ml final concentration), yeast tRNA (0.2 mg/ml), and poly$(dA)_{40\text{-}60}$ (0.4 mg/ml). Hybridization was performed on a Ventana Discovery system for 6 hr at 42C (Ventana Medical Systems, Tucson, AZ). Microarrays are washed to a final stringency of 0.1X SSC. Microarrays were scanned on a dual-channel, dynamic autofocus, fluorescent scanner at 10 um resolution (Agilent Technologies, Palo Alto, CA). Fluorescent intensity was determined using Imagene™ software (BioDiscovery, Marina del Rey, CA).

Normalization and Scaling of microarray data. Signals from independent samples can vary on a global-basis and must be adjusted to a common standard for comparison. Adjustment of expression levels in compared samples was performed as previously described (21). Briefly, compared samples were first normalized using low level noise signals (commonly referred to as additive noise (AN) (74). The parameters of the AN were calculated from nonexpressed genes whose signal values exhibited a normal distribution. The mean and standard deviation (SD) of the AN signals was obtained by nonlinear curve fitting after exclusion of expressed genes from the distribution. Expression values from a given chip were then normalized such that the AN

distribution had a mean of 0 and a SD of 1. Genes expressed 3 SD above the mean of AN are defined as expressed genes and used for further analysis. A second scaling step is then performed on expressed genes that were scaled to a common standard through a robust linear regression analysis.

Selection of hypervariable genes (HV-genes). Genes responsive to CSCs were identified using an analysis of temporally induced gene expression changes. This procedure utilizes an internal standard, denoted "the reference group" to define the levels of technologic and normal biologic variance in the experiment so that these values can be used to define stimuli-induced variation in a statistically robust manner. The majority of genes in the control group are not sensitive to temporal changes. The reference group is therefore composed of a group of genes statistically significantly expressed above the mean of AN in control samples, whose residuals approximate a normal distribution based on the Kolmogorov-Smirnov criterion, and that have low variability of expression over time as determined by an F-test. Variance in the reference group is due only to technical variation and normal biologic variation and therefore the distribution of expression of the reference group can be used to identify genes that vary due to experimental conditions in a manner that is statistically significantly higher than the technologic and normal biologic variance of the system using an F-test. Genes identified using these procedures are denoted "hypervariable genes".

F-means cluster analysis of HV-genes co-expression. Groups of genes that vary in expression over time in a similar manner, based on the technologic and normal biologic variation in the system, are included in a given cluster. The reference group defined above is once again used as a reference to define statistically significant thresholds for clustering parameters used in an F-

test. In this manner, the variance of the system is used to define the number of clusters thus removing the subjective nature of most clustering methods. The method is not without some subjective criterion as genes can belong to multiple clusters. In this method, a given gene is placed into the largest cluster such that the broadest biologic phenomena of the system, that is those involving the largest number of genes, can be distinguished. To do this, clustering is begun by defining a simple parameter for each HV gene. This parameter, denoted connectivity, is equal to the number of genes that vary in expression in a similar manner as a given gene. Clusters are nucleated starting with genes of highest connectivity. Genes of lower connectivity will be included in a given cluster if their expression varies over time in a manner similar to the gene used to nucleate the cluster, i.e. if their deviations of expression over time do not exceed the variation of the residuals in the reference group based on an F test.

Correlation Coefficient Analysis. F-clustering was used to identify the kinetic behavior of genes for each stimulus. Correlation coefficient analysis was used to identify genes that behave in a similar manner among groups. In this analysis, a Pearson correlation coefficient is used for clustering of genes with similar time-dependent behavior among groups. A correlation threshold was established using a Monte-Carlo simulation experiment such that the chances of identifying a false positive or false negative selection is <1. Matrices of correlation coefficients are calculated for these clusters and are represented in a graphical output termed a connectivity mosaic such that patterns of correlated and non-correlated behavior of genes can be identified by visual inspection.

Discriminant function analysis (DFA). DFA is a method that identifies a subset of genes whose expression values can be linearly combined in an equation, denoted a root, whose overall value is

distinct for a given characterized group. DFA therefore, allows the genes that maximally discriminate among the distinct groups analyzed to be identified (61). In the present work, a variant of the classical DFA, named the Forward Stepwise Analysis, was used for selection of the set of genes whose expression maximally discriminates among experimentally distinct groups. The Forward Stepwise Analysis was built systematically. Specifically, at each step all variables were reviewed to identify the one that most contributes to the discrimination between groups. This variable was included in the model, and the process proceeds to the next step. The statistical significance of discriminative power of each gene was also characterized by partial Wilk's Lambda coefficients (15), which are equivalent, to the partial correlation coefficients generated by multiple regression analyses. The Wilk's Lambda coefficient used a ratio of within group differences and the sum of within plus between group differences. Its value ranged from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power).

RESULTS

Gene Expression Alterations Induced by Cigarette Smoke Condensates (CSCs). In order to determine if CSCs modulate cellular physiology pleiotropically and in a temporally complex manner, monolayer cultures of NHBE cells were treated in logarithmic phase of growth for up to 12 hours with CSC-A or CSC-B in the presence of 2% S9 metabolic fraction, or with 2% S9 fraction alone. Cell viability after 12 hours exposure was 84% and 73% for CSC-A and CSC-B treatments, respectively, when compared to untreated cells. RNA was extracted from cells at 2, 4, 8, and 12 hours post-treatment, fluorescently labeled and hybridized to genome-scale microarrays. CSC-induced changes in gene expression were determined in a comprehensive manner using a recently described method of analysis, denoted hypervariable analysis, which is based on the observation that gene expression for a majority of genes is relatively stable among replicates in untreated cells. Any measurable variation in this large set of genes by microarray analysis reflects the combined effects of intrinsic normal biologic variation and extrinsic technological variation in an unmanipulated cell. Genes that are impacted by exposure to CSCs, and whose mRNA expression varies over time in a statistically significant manner that is greater than this normal biologic and technical variation, are termed "hypervariable" (HV).

Of the 21,349 genes and open reading frames (ORFs) on the high-density array used in this experiment, a combined total of 4,894 (22.9%) were classified as HV after CSC treatment. Individually, the expression of 3,665 genes/ORFs (i. e., 17.2% of all the genes/ORFs on the array), and 3,668 genes/ORFs (17.2%) were hypervariable in at least one time point during the 12-hour exposure period to CSC-A and CSC-B respectively (Fig. 1A, Online Supplemental Table S1). The observation that the expression of a large number of genes is altered in a significant manner during the 12 h treatment demonstrates a significant impact by CSCs on steady-state levels of mRNAs in NHBE cells. A majority of the HV genes (i.e. 2,439) were

common to both CSC-treated groups, suggesting that the two CSCs affected cells largely in a similar manner. However, unique non-overlapping sets of HV genes were also identified after treatment with CSC-A (i.e., 1226 genes) and CSC-B (i.e., 1229 genes), which may reflect specific quantitative and/or qualitative differences in the various classes of chemical constituents comprising the two CSCs.

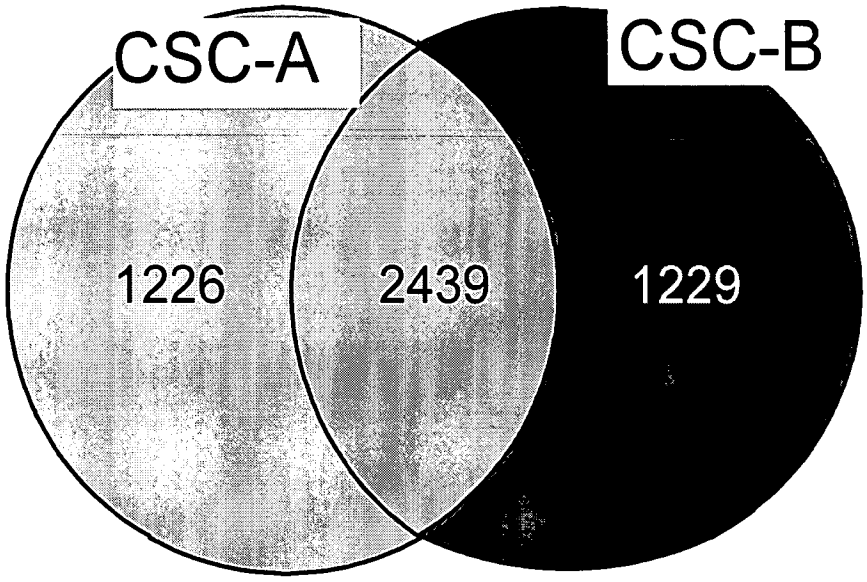


Figure 1A: Venn diagram comparing gene expression modulations induced by CSC-A (3665) and CSC-B (3668). The number of genes affected is given in each sector. The intersections between sectors reflect the number of genes that are affected by both CSCs (2439).

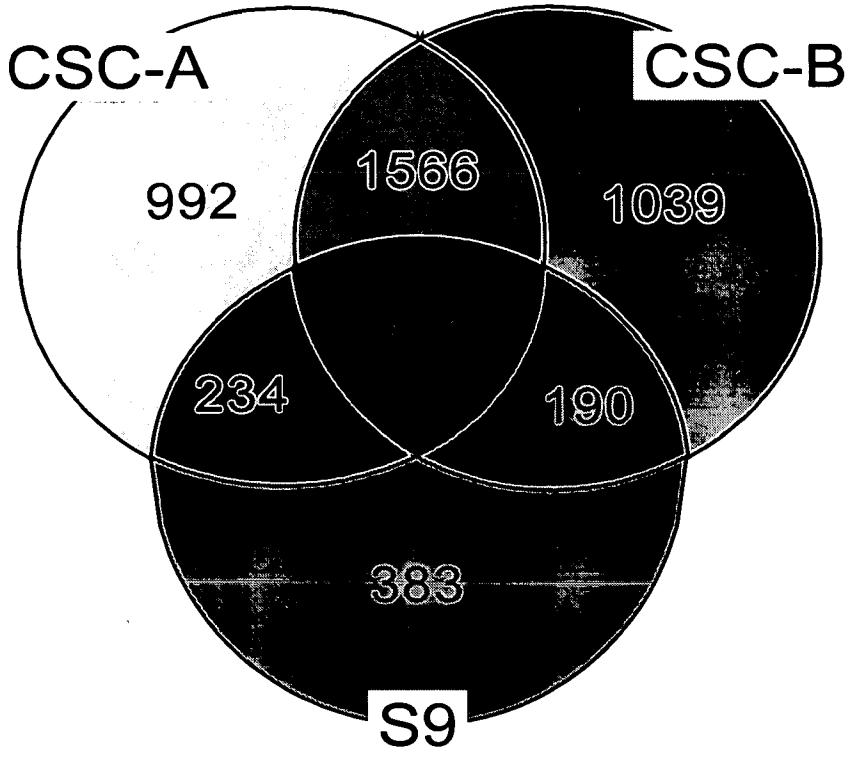


Figure 1B: Venn diagram comparing gene expression modulations induced by CSC-A (3665), CSC-B (3668), and S9 metabolic fraction (1680). The number of genes affected by each treatment is given and the intersections between sectors reflect the number of genes that are affected by more than one treatment (e.g., a common set of 873 genes is affected by CSC-A, CSC-B and S9).

Subsequent to exposure in vivo, the human body attempts to detoxify, neutralize, and eliminate cigarette smoke toxins through the action of Phase I and Phase II enzymes functioning in various metabolic pathways (34; 39). However, during this detoxification process a number of procarcinogenic compounds in cigarette smoke are bioactivated into reactive electrophiles that have potent carcinogenic potential in exposed cells. Thus, in order to dissect the full biological potential of complex chemical mixtures such as CSCs, it is standard procedure for in vitro

studies to include co-treatment with a S9 microsomal fraction from Aroclor 1254-treated rats, which provides the appropriate enzymes that mimic the detoxification process in mammalian cells. Similarly, in the present studies, NHBE cells were exposed to CSC in conjunction with S9. Consequently, as an important control it was necessary to discriminate the effects of S9 alone on gene expression. Therefore, we performed a HV analysis on microarray results from cells treated only with S9 for 2, 4, 8, and 12 hours. Several interesting observations emerged from this analysis. First, we noted that the expression of 1680 (7.9%) genes became HV sometime during the 12-hour exposure period with S9 (see Figure 1B and Online Supplemental Table S1). Second, Figure 1B also shows that 1297 of these 1680 genes were also HV in one or both CSC treatments, which is not surprising since all three treatment conditions (i.e., CSC-A, CSC-B, and S9) had the same concentration of S9. Third, as we show below, even though the CSCs and S9 induce a HV state in a large common set of genes, CSCs and S9 do not affect these genes in similar ways indicating differential kinetic effects between S9 alone and S9 in context with CSCs.

Gene Expression Kinetics. Subsequent to determining that the complex mixture of toxins and carcinogens in CSCs has a broad impact on the transcriptome of NHBE cells, we hypothesized that sustained treatment over a 12-hour period would also allow detection, not only of alterations such as induction and suppression, but of gene induction/suppression with transient, sustained, or periodic characteristics. To test this idea, we defined the kinetic effects of gene expression profiles generated from cells treated with CSC-A, CSC-B, or S9 from 0-12 hours using F-cluster analysis, which is a statistically robust method for defining clusters of genes with similar expression patterns over time. In this analysis, the normal variance of the system is calculated and used to identify a statistical threshold for cluster selection at which groups of genes are likely

to cluster by chance. This threshold is then used for further analysis to ensure the statistical robustness of the clustering process. The biologic significance of the cluster is related to cluster size, as the largest clusters identified represent synchronous changes in the greatest number of cellular processes (80). Specifically, larger clusters represent, in a statistically robust manner, the most significant experimentally induced processes in these cells. When F-cluster analysis was applied to the total HV set of 4894 genes/ORFs, 306 clusters were defined by statistical analysis, the majority of which contained less than 50 member genes. Cluster numbers were arbitrarily assigned from -150 to 150, with the corresponding positive and negative numbers representing complementary gene expression patterns (e.g. steady increase in expression over time compared to a steady decrease in expression).

In each of the three treatment conditions clusters containing 50 or more genes were chosen for further characterization because this cutoff generated a sufficient number of large clusters that adequately represented the major kinetic changes caused by each treatment (see Fig. 2 A-C and Online Supplemental Table S2). As predicted, gene expression changes induced by CSCs were complex, with the majority of clusters in CSC-treated cells being multi-modal (see Figures 2A and B). For example, in CSC-A-treated cells, genes in clusters 1, 3, 7, 12, 15, and 22 were up-regulated within the first two hours, began to return to baseline, then were once again induced late in the experiment, suggesting initial treatment effected gene expression and some secondary effect, e.g. a CSC metabolite or the action of early gene expression changes, reinitiated a cellular response. (Fig. 2A). While genes within each of these clusters show early increases in expression (within the first 2 h of treatment), suggesting CSC-A treatment has immediate effects on cells, Clusters 18, 30, 35, and 39 show a later increase in gene expression (i.e., $\geq$ 4h). Figure 2B shows that in CSC-B treated cells cluster analysis shows that gene expression peaks primarily between 4-8 hours, as opposed to a 2 hour peak in CSC-A treated

cells, suggesting that some of the effects of CSC-B treatment are delayed with respect to those of CSC-A (e.g., see clusters 4, 5, 9, 10, 16, and 32*)*. These data are in distinct contrast to the major clusters of genes in S9-only treated cells, which displayed simple kinetics, i.e., expression decreasing or increasing continuously over time (Figure 2C). Although 66% of HV genes affected by CSC-A and CSC-B were identical (see Figure 1), it is clear from Figure 2 that the expression kinetics for these genes were nevertheless distinct for the two CSCs. This is evidenced by the fact that the predominant coordinated behavior in CSC-A-treated cells is represented by the largest cluster (i.e., cluster 1), that contains 1063 HV genes and whose expression peaked at 2 hours post-treatment. This is in contrast to CSC-B-treated cells in which the predominant behavior of genes is represented by cluster 2, which contains 1,036 genes and whose expression peaked at 4-8 hours, suggesting that some of the effects of CSC-B treatment are delayed with respect to those of CSC-A.

Figure 2A: ***Clusters containing 50 or more genes in CSC-A-treated cells.***

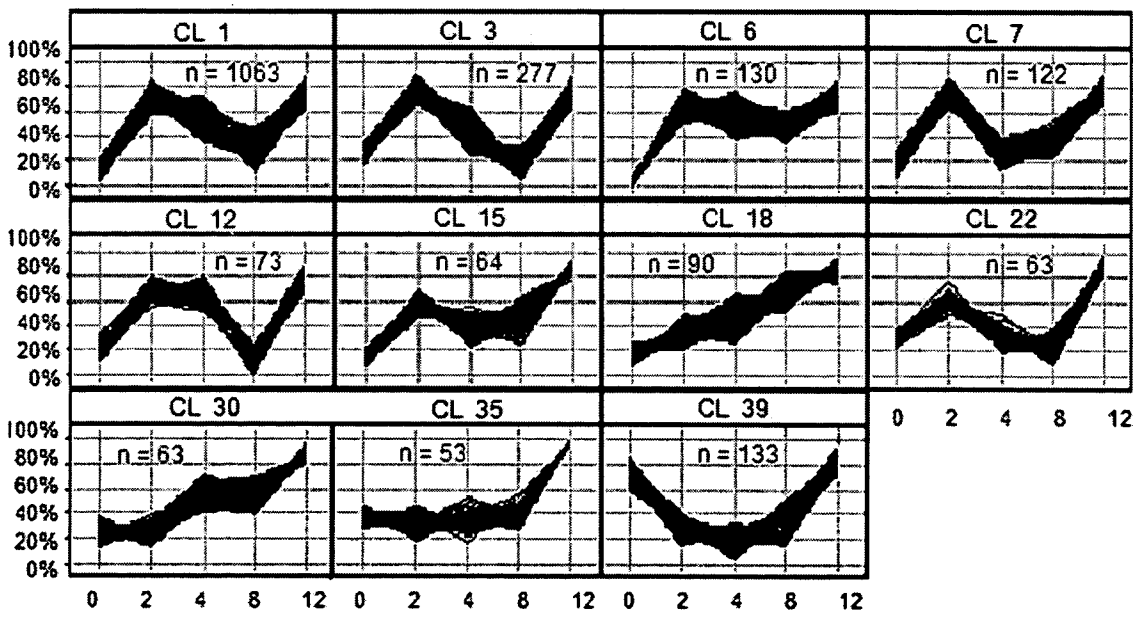


Figure 2B: ***Clusters containing 50 or more genes in CSC-B-treated cells.***

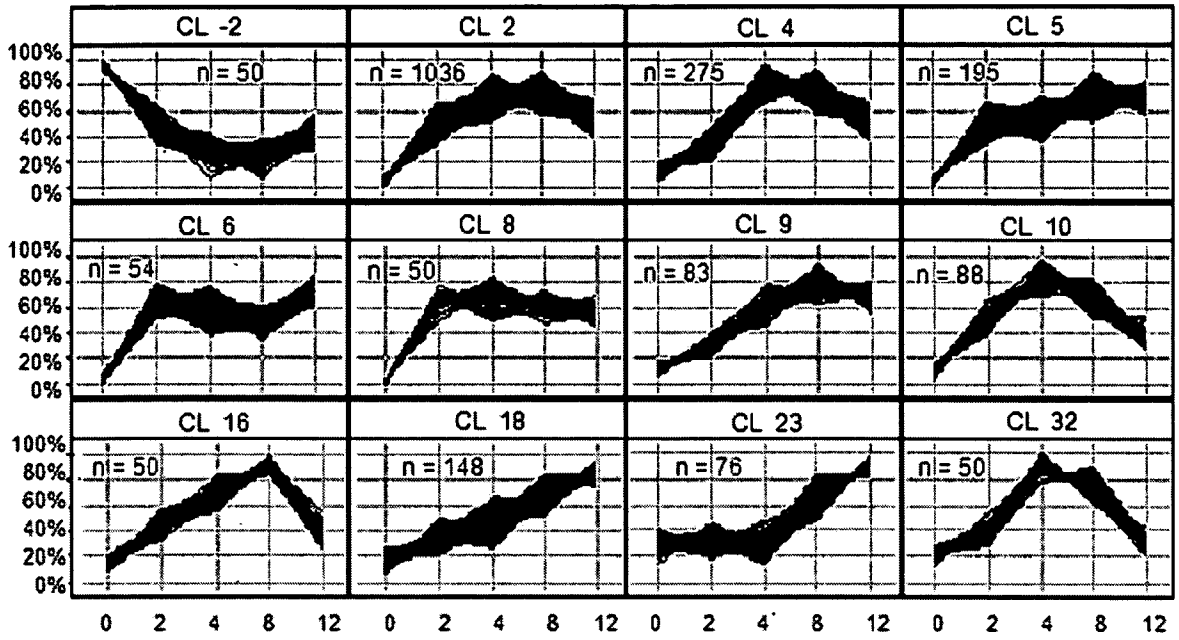
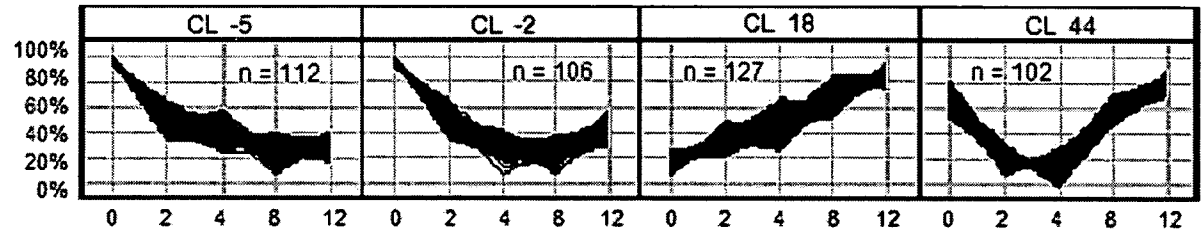


Figure 2C: ***Clusters containing 50 or more genes in S9-treated cells.***



Figures 2A-2C: F-clusters of genes containing 50 or more members. Gene expression profiles between 0 and 12h are expressed a percent of highest expression value for each gene. F-cluster numbers are given at the top of each cluster of profiles. The number of member genes in each cluster (n) is shown in red for each cluster.

Since clusters with a large number of member genes reflect predominant biological behavior patterns that are likely to be functionally interrelated, we hypothesized that the cluster 1

set of 1063 genes from CSC-A-treated cells and the cluster 2 set of 1036 genes from CSC-B-treated cells corresponded to important biological phenomena common to the two CSCs. If this speculation is correct, then despite the fact that CSC-A and CSC-B treatments modulate genes in a temporally distinct manner, the two clusters should contain many of the same genes. We found, in fact, that a set of 554 genes (approximately 50% of the genes in each cluster) are present in both cluster 1 (from CSC-A) and cluster 2 (from CSC-B). A total of 330 genes from this set of 554 genes (59.5%) have known functions while the remaining 224 are ORFs (see Online Supplemental Table 1S). Functional classification of these 330 genes common to cluster 1 and cluster 2 indicates that 10% have functional roles in proliferation, 12.4% in transcription, 4.5% in apoptosis, and 5.1% in damage/repair responses. In addition, as shown in Table 1 below, 34 (10%) of these genes are documented as having potential roles in several major diseases caused by long-term tobacco exposure, i.e., lung cancer, coronary heart disease, and asthma.

Table 1: Genes Common to Clusters 1 and 2 with roles in tobacco-related diseases

GenBank accession no.	Gene Abbreviation	Gene description	Disease
NM_001613	ACTA2	Actin, alpha 2, smooth muscle, aorta	Lung Cancer
NM_005181	CA3	Carbonic anhydrase III, muscle specific	Lung Cancer
NM_005199	CHRNG	Cholinergic receptor, nicotinic, gamma polypeptide	Lung Cancer
NM_002594	PCSK2	Proprotein convertase subtilisin/kexin type 2 (PC2)	Lung Cancer
NM_004624	VIPR1	Vasoactive intestinal peptide receptor 1 (VPAC1)	Lung Cancer
NM_004448	ERBB2 (HER2/NEU)	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2	Lung Cancer
NM_024083	ASPSCR1	Alveolar soft part sarcoma chromosome region, candidate 1	Lung Cancer

NM_003872	NRP2	Neuropilin 2	Lung Cancer
U33749	TITF1	Thyroid transcription factor 1	Lung Cancer
NM_002639	SERPINB5	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5, (maspin)	Lung Cancer
AF135794	AKT3	V-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)	Lung Cancer
NM_001618	ADPRT	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase) PARP1	Lung Cancer
NM_016434	TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	Lung Cancer
NM_003072	SMARCA4 (BRG1)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	Lung Cancer
NM_004061	CDH12	Cadherin 12, type 2 (N-cadherin 2)	Lung Cancer
U28749	HMGIC	High-mobility group (nonhistone chromosomal) protein isoform I-C	Lung Cancer
NM_002592	PCNA	Proliferating cell nuclear antigen	Lung Cancer
NM_033215	PPP1R3F	Protein phosphatase 1, regulatory (inhibitor) subunit 3F (PPP1R3F), mRNA	Lung Cancer
NM_006218	PIK3CA	Phosphoinositide 3-kinase, catalytic, alpha polypeptide	Lung Cancer
NM_005506	CD36L2	CD36 antigen (collagen type I receptor, thrombospondin receptor)-like 2 (lysosomal integral membrane	Lung Cancer
NM_004994	MMP9	Matrix metalloproteinase 9	Lung Cancer
NM_003810	TNFSF10	Tumor necrosis factor (ligand) superfamily, member 10 (TRAIL)	Lung Cancer
NM_002961	S100A4	S100 calcium binding protein A4 (calcium protein, calvasculin, metastasin, murine placental homolog)	Lung Cancer
NM_007084	SOX21	SRY (sex determining region Y)-box 21	Lung Cancer
NM_003682	MADD	MAP-kinase activating death domain (DENN)	Lung Cancer
BC002712	MYCN	V-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	Lung Cancer

NM_004353	SERPINH1	Serine (or cysteine) proteinase inhibitor, clade H), member 1, HSP47	Oral Cancer
NM_000640	IL13RA2	Interleukin 13 receptor, alpha 2	Asthma
NM_002046	GAPD	Glyceraldehyde-3-phosphate dehydrogenase	Asthma
NM_021804	ACE2	Angiotensin I converting enzyme (peptidyl-dipeptidase A) 2	Coronary Heart Disease
NM_017614	BHMT2	Betaine-homocysteine methyltransferase 2	Coronary Heart Disease
NM_020974	CEGP1	CEGP1 protein	Coronary Heart Disease
NM_018641	C4S0	Chondroitin 4-O-sulfotransferase 2	Coronary Heart Disease
NM_006874	ELF2	E74-like factor 2 (ets domain transcription factor), NERF	Coronary Heart Disease

In clear contrast to both CSC-A and CSC-B, the S9 treated cells show a pronounced tendency towards suppression of gene expression. An F-clustering analysis of the S9 data (shown in Figure 2C) resulted in only four clusters that contained 50 or more genes. Clusters 2, 5, and 44 all show decreases in gene expression level with a nadir at 4-8h. Cluster 18 contains genes that show an increase in gene expression levels, but whose expression peaks at 12h, which is notably different from the robust early gene responses elicited by treatment with both CSCs. Additional evidence that the overall effects of S9 and CSCs on gene expression levels are quite distinct is evident when traditional hierarchical clustering algorithms are used to compare the overall differences in HV gene expression in each treatment group over the entire 12-hour time course. Figure 3 shows the results of this analysis for the common subset of genes that were HV in all three treatment groups (i.e., the 873 genes denoted in Figure 1). The major observation is that the expression data for these 873 genes partition into two separate groups with S9-treated cells being clearly distinguishable from CSC-A and CSC-B treated cells, which are similar to each other. The data further indicate that S9 exerts a largely suppressive effect on the transcriptome of NHBE cells in contrast to a predominant inductive effect of CSC-A and CSC-B.

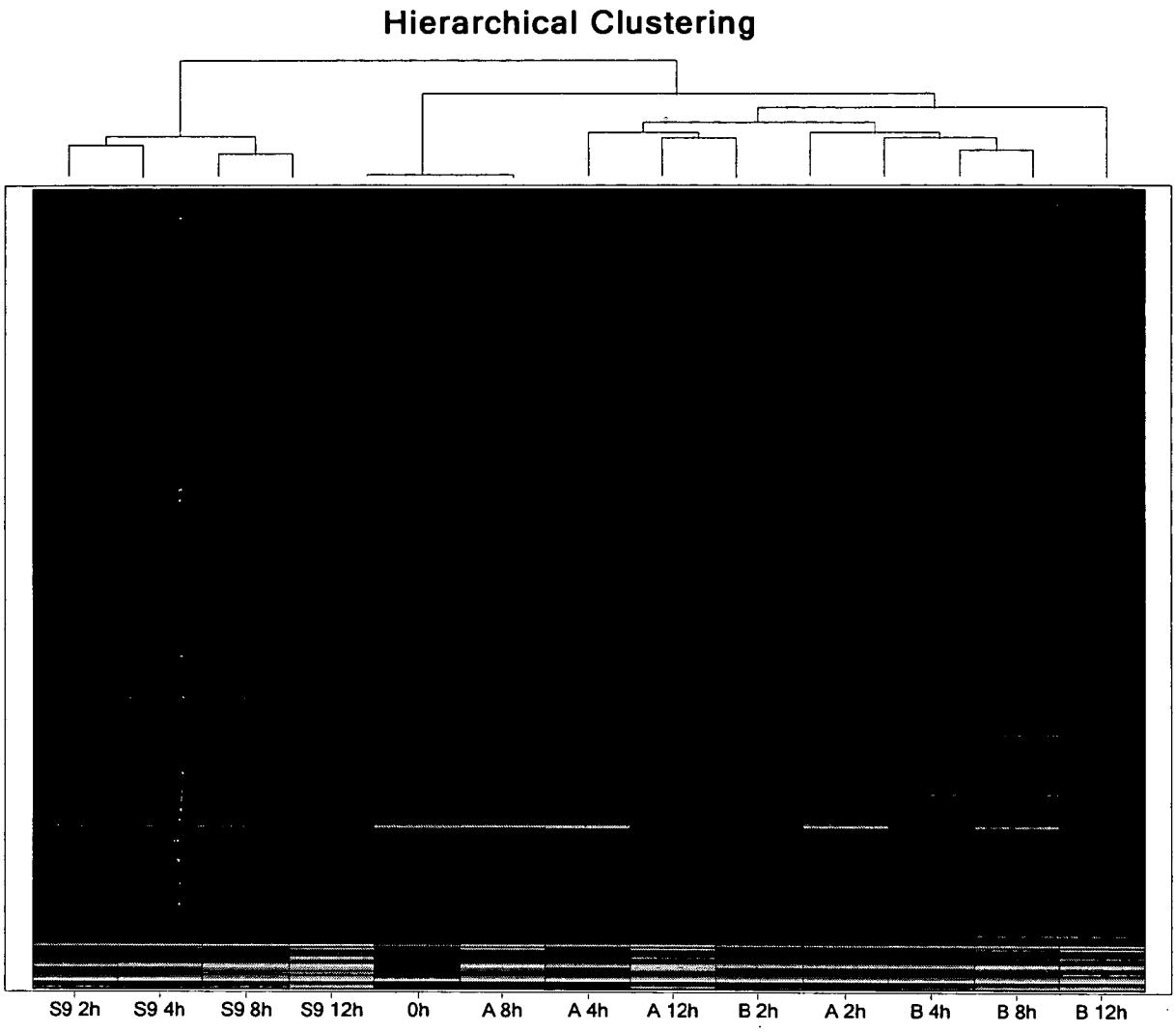


Figure 3. Cluster analysis of genes that were HV in all three treatment groups (A: CSC-A, B: CSC-B, S9). Dendrogram depicts the hierarchical relationship between the three treatments based on their gene expression patterns at all time points from 0 – 12 hours.

0 Log_2 3.0

Defining CSC-specific Toxicological Effects. As shown in Figures 2 and 3, CSCs induce a range of temporally distinct alterations to the homeostatic transcriptome of the NHBE cell that are unique in that they are qualitatively and quantitatively dissimilar from the effects of S9. In an attempt to define a biological context for these data, we used correlation analyses to identify genes whose expression changes were highly correlated in CSC-A and CSC-B treated cells but not in S9-treated cells. This was achieved using a Monte Carlo analysis to establish a statistical threshold above which correlated behavior was unlikely to have occurred by chance. In this approach, gene expression levels are randomized maintaining the same mean and standard deviation. A correlation coefficient is then identified above which no genes are correlated in the randomized data sets. The probability that genes that correlate in experimental data sets above this threshold will occur by chance is <1/total number of genes analyzed. As shown in Table 2, this method identified 40 HV genes whose expression changes were correlated in CSC-A and CSC-B treated cells but not in S9-treated cells. The similarities between the two tobacco-treated sample groups can be visualized by applying correlation coefficient analysis to the genes within a given treatment, representing this visually in a correlation mosaic, and comparing the visual pattern of the mosaic to other such mosaics generated using data from different treatments. The correlation coefficients of these genes are presented in a correlation mosaic color map (see Figure 4) in which genes with highly correlated behavior are denoted by a red pixel, and genes with highly negatively correlated behavior by a blue pixel. This mosaic provides a means of assessing the similarities of expression behavior of the correlated genes in CSC-A, CSC-B, and S9-treated cells by visual inspection.

Table 2: HV Genes Specific for CSC-A and CSC-B Treatment

GenBank	Gene	Gene description

accession no.	abbreviation	
AB032985	NXPH3	Neurexophilin 3
AB046848	KIAA1628	KIAA1628 protein
AB058772	SEMA6C	Sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6C
AF178532	BACE2	Beta-site APP-cleaving enzyme 2
BC015737		Homo sapiens, ninjurin 2, clone MGC:22993 IMAGE:4907813
BC015929	NR1D2	Nuclear receptor subfamily 1, group D, member 2
BC017732	STRBP	Spermatid perinuclear RNA binding protein
M23326	TRDV3	T cell receptor delta variable 3
NM_000341	SLC3A1	Solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine), member 1
NM_000663	ABAT	4-aminobutyrate aminotransferase
NM_000922	PDE3B	Phosphodiesterase 3B, cGMP-inhibited
NM_000981	RPL19	Ribosomal protein L19
NM_001383	DPH2L1	Diptheria toxin resistance protein required for diphthamide biosynthesis-like 1 (S. cerevisiae)
NM_002046	GAPD	Glyceraldehyde-3-phosphate dehydrogenase
NM_002757	MAP2K5	Mitogen-activated protein kinase kinase 5
NM_002890	RASA1	RAS p21 protein activator (GTPase activating protein) 1
NM_003286	TOP1	Topoisomerase (DNA) I
NM_003408	ZFP37	Zinc finger protein 37 homolog (mouse)
NM_004057	CALB3	Calbindin 3, (vitamin D-dependent calcium binding protein)
NM_004066	CETN1	Centrin, EF-hand protein, 1
NM_004083	DDIT3	DNA-damage-inducible transcript 3
NM_004282	BAG2	BCL2-associated athanogene 2
NM_004846	EIF4EL3	Eukaryotic translation initiation factor 4E-like 3
NM_004939	DDX1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1
NM_005476	GNE	UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase
NM_005619	RTN2	Reticulon 2
NM_007217	PDCD10	Programmed cell death 10

NM_007275	FUS1	Lung cancer candidate
NM_012192	FXC1	Fracture callus 1 homolog (rat)
NM_012288	KIAA0057	TRAM-like protein
NM_013366	APC2	Anaphase-promoting complex subunit 2
NM_013401	RAB3IL1	RAB3A interacting protein (rabin3)-like 1
NM_014395	DAPP1	Dual adaptor of phosphotyrosine and 3-phosphoinositides
NM_015057	KIAA0916	KIAA0916 protein
NM_017491	WDR1	WD repeat domain 1
NM_017581	CHRNA9	Cholinergic receptor, nicotinic, alpha polypeptide 9
NM_020122	PCMF	Potassium channel modulatory factor
NM_020685	HT021	HT021
NM_021120	DLG3	Discs, large (Drosophila) homolog 3 (neuroendocrine-dlg)
NM_031310	PLVAP	Plasmalemma vesicle associated protein

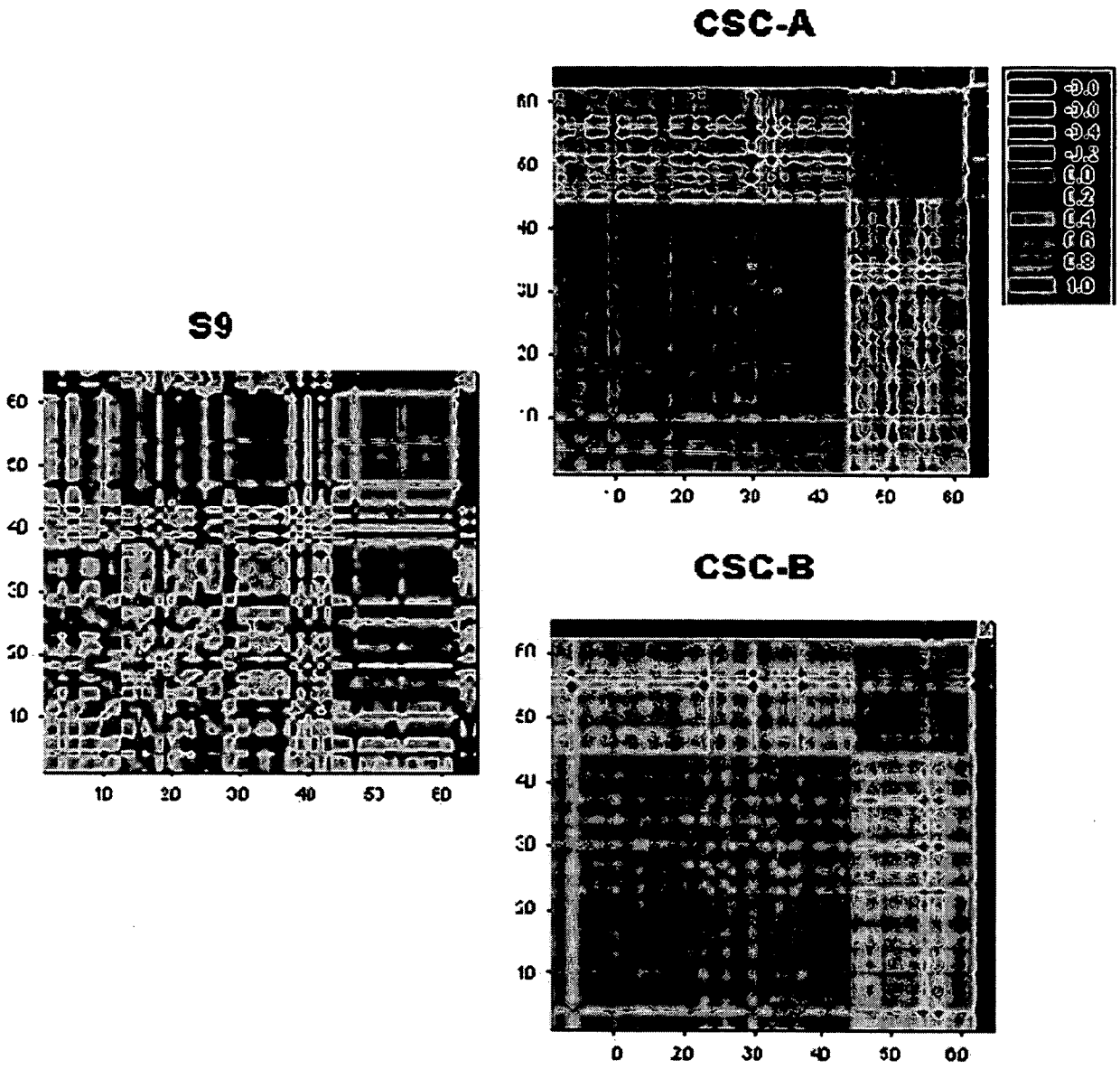


Figure 4. Correlation mosaics of genes listed in Table 2. Correlation coefficients were generated for each of the 40 genes in Table 2, comparing the set to itself in each of the three conditions. The same gene order runs across the x and y axes of the

mosaics. Correlation mosaics for HV genes highly correlated in response to CSC-A and CSC-B, and not correlated with responses to S9. Each pixel in the plot represents a correlation coefficient of gene expression. Genes highly positively correlated are denoted in red and those highly negatively correlated are in blue. The same order of the genes along axis is used for all three mosaics. Genes highly correlated in CSC-A and CSC-B, but not in S9-treated cells are denoted as a red cluster in the lower left hand corner of CSC-A and the CSC-B mosaic. This cluster is disrupted in the S9 mosaic demonstrating the variance in gene regulation that occurred in S9-treated cells.

The highly correlated expression characteristics of the CSC-impacted genes identified by this analysis suggest that these genes are likely to participate in pathways relevant to the effects specific to CSCs and not to S9. We attempted to define these pathways using PathwayAssist™ software (Stratagene, La Jolla, CA), a commercially available visualization engine that scans and assesses documented literature and available standardized databases in order to filter, classify, and prioritize proteins in terms of their functional relationships to known biological pathways. The results, portrayed in Figure 5, highlight the fact that this set of genes encodes proteins that play key roles in pathways that are relevant to the documented pathological effects of cigarette smoke. For example, several of the genes listed in Table 2 are implicated in lung oncogenesis (i.e., FUS1, GAPD, & semaphorin), in various types of dysfunctions in lung cells involving apoptosis (i.e., PDE3B, PDCD10), in cell cycle control (MAP2K5, RASA1, APC2, RASA1), in DNA topology and DNA repair (TOP1, DDIT3), and in cellular stress (BAG2). In addition, several genes are involved in neurosignaling (neurexophilin, KIAA1628), neuroregeneration (semaphorin), neuropathology (BACE2, ABAT, DLG3), and inflammation (NINJ2, TRDV3,

SLC3A1). The induction of a range of neuroendocrine-related genes is interesting in light of the fact that many small cell lung cancers and some non-small cell lung cancers exhibit a variety of pathological and molecular features of pulmonary endocrine cells, and can be stimulated by an autocrine/paracrine array of neuroendocrine peptides (9). Accordingly, expression of neuroendocrine markers has been shown to be useful in the differential diagnosis of lung cancers (56). The gene set shown in Table 2 also includes CHRNA9, a human nicotinic acetylcholine receptor expressed in several tissues including inner ear hair cells, brain, and in activated fibrosarcoma cells and whose relevance to nicotine signaling in primary lung cells is as yet uncharacterized (37; 55).

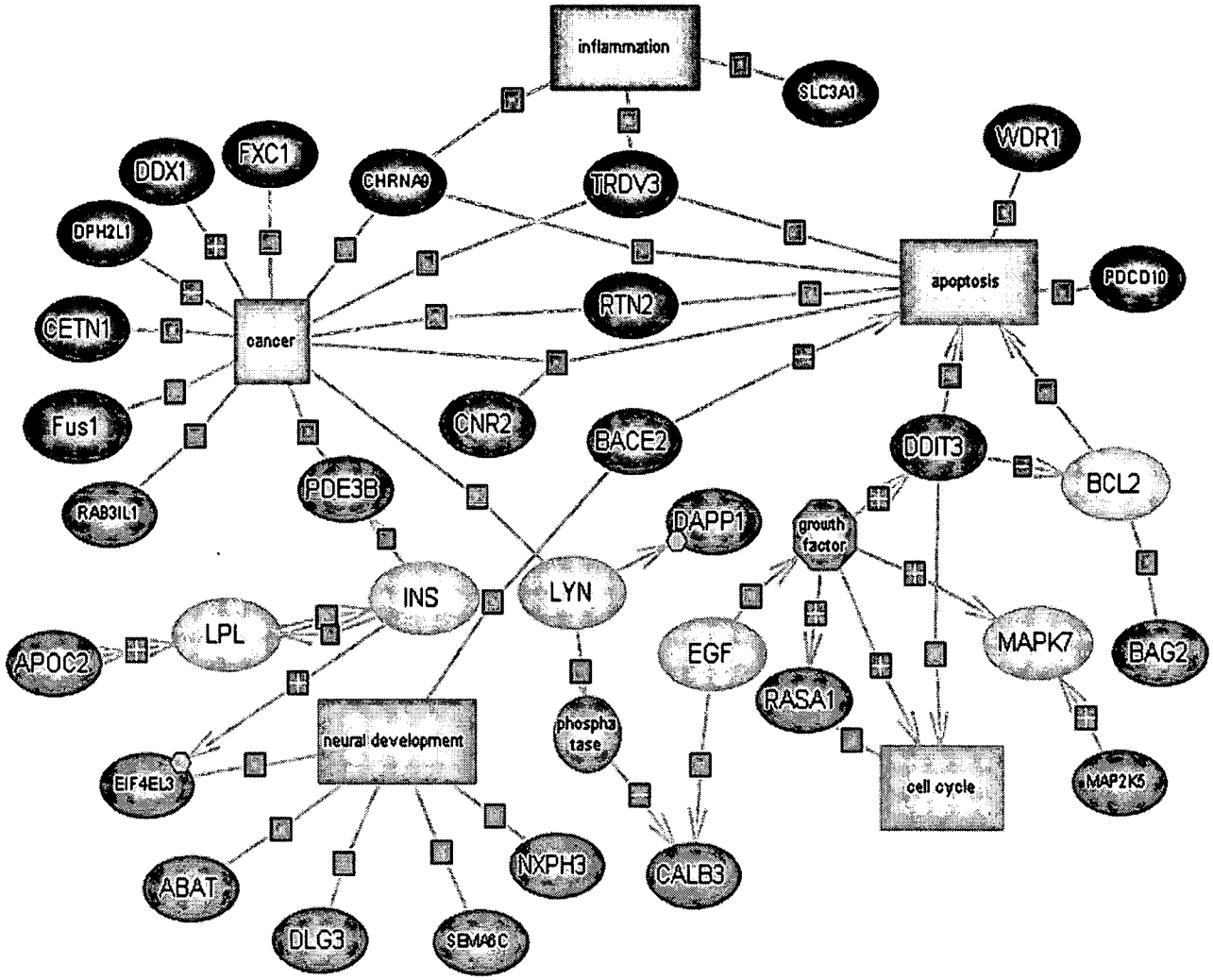


Figure 5. Functional associations of HV genes specific for CSC-A and CSC-B treatment. The expression patterns of this set of genes are highly correlated in CSC-treated NHBE cells and not correlated with those seen in cells treated with S9 alone. Red ovals indicate genes from Table 2. Grey ovals (indicating additional proteins not in Table 2) were added to better define the regulatory networks of the genes identified in this analysis. Orange ovals indicate classes of functional peptides. Yellow rectangles indicate cellular processes in which these genes participate. Each line indicates a regulatory relationship (binding,

regulation, etc.) based upon a literature reference. Regulatory relationships are denoted in a box on the line with positive regulation represented as a plus sign, negative regulation as a minus sign, and unknown relationships by no sign

Defining S9-Specific Effects. Using a similar analysis as described for CSCs in Table 2 and Figure 4, we assessed the global effects of S9 by first identifying the subset of HV genes that are correlated among all three treatment groups and then assuming that the effect on these genes is due to S9 solely, since their expression characteristics did not change when S9 was combined with a CSC. As described above, we performed a Monte Carlo analysis to define a statistically robust correlation coefficient unlikely to occur by chance. Using this threshold, the probability of identifying a gene correlated in all three groups by chance is <1/total number of genes analyzed, thereby confirming the high statistical specificity of this method. As shown in Table 3, a set of 52 genes was identified and the probable function of these genes was assessed using PathwayAssist™ software (Fig. 6). Many of the genes appear to have roles in modulating apoptosis (e.g., AVEN, LIG1, PTEN, etc.) suggesting that the predominant cellular response to chronic S9 exposure is to activate apoptotic programs (12; 69). A second group of S9-modulated genes modulates cellular surface chemistry, adhesion, and cellular differentiation (e.g., SIAT4B, KRT10, CDSN and EXT2) (11; 31; 44). These results suggest that the standard practice of including S9 in toxicogenetic experiments significantly modulates cellular physiology, which may complicate and bias the results assessing the effects of CSCs or any other type of complex hydrocarbon mix requiring metabolic activation.

Table 3: Genes Specific for S9 Treatment

GenBank accession no.	Gene abbreviation	Gene description
NM_001303	COX10	COX10 homolog, cytochrome c oxidase assembly protein
AK056540		Homo sapiens cDNA FLJ31978, weakly similar to Probable hexosyltransferase
NM_016013	LOC51103	CGI-65 protein
NM_031916	ASP	AKAP-associated sperm protein
NM_000947	PRIM2A	Primase, polypeptide 2A (58kD)
NM_006927	SIAT4B	Sialyltransferase 4B
NM_006441	MTHFS	5,10-methenyltetrahydrofolate synthetase
NM_002699	POU3F1	POU domain, class 3, transcription factor 1
NM_002954	RPS27A	Ribosomal protein S27a
AK055508	FLJ11785	Rad50-interacting protein 1
NM_024636	FLJ23153	Likely ortholog of mouse tumor necrosis-alpha-induced adipose-related protein
BC011231		Homo sapiens, Similar to angiotensinogen
NM_007052	NOX1	NADPH oxidase 1
NM_000234	LIG1	Ligase I, DNA, ATP-dependent
NM_032553	FKSG79	Putative purinergic receptor
NM_000025	ADRB3	Adrenergic, beta-3-, receptor
AF023203		Homo sapiens homeobox protein Og12
U50536		Human BRCA2 region, mRNA sequence CG011
NM_000421	KRT10	Keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)
NM_001264	CDSN	Corneodesmosin
NM_000355	TCN2	Transcobalamin II; macrocytic anemia
NM_000401	EXT2	Exostoses (multiple) 2

NM_014214	IMPA2	Inositol(myo)-1(or 4)-monophosphatase 2
NM_003797	EED	Embryonic ectoderm development
AF319523		Homo sapiens RT-LI mRNA, complete sequence
AF074331	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2
AF189011	RNASE3L	Putative ribonuclease III
BC009752		Homo sapiens, Similar to sex comb on midleg-like 1 (Drosophila)
NM_000691	ALDH3A1	Aldehyde dehydrogenase 3 family, memberA1
NM_006006	ZNF145	Zinc finger protein 145 (expressed in promyelocytic leukemia)
NM_005831	NDP52	Nuclear domain 10 protein
L26584	RASGRF1	Ras protein-specific guanine nucleotide-releasing factor 1
NM_014182	HSPC160	HSPC160 protein
NM_004963	GUCY2C	Guanylate cyclase 2C (heat stable enterotoxin receptor)
AB023223	STXBP-TOM	Tomosyn
NM_018919	PCDHGA6	Protocadherin gamma subfamily A, 6
NM_002968	SALL1	Sal-like 1 (Drosophila)
NM_003587	DDX16	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
AK024449	PP2135	PP2135 protein
AB034205	LUC7A	Cisplatin resistance-associated overexpressed protein
BC011589	OSM	Oncostatin M
NM_006597	HSPA8	Heat shock 70kD protein 8
NM_004384	CSNK1G3	Casein kinase 1, gamma 3
AK057672		Homo sapiens cDNA FLJ33110 fis
NM_016344	PRO1900	PRO1900 protein
NM_018651	ZFP	Zinc finger protein
NM_004717	DGKI	Diacylglycerol kinase, iota
NM_006479	PIR51	RAD51-interacting protein
AK024250		Homo sapiens cDNA FLJ14188 fis

NM_001382	DPAGT1	Dolichyl-phosphate N-acetylglucosaminephosphotransferase 1
NM_020371	AVEN	Cell death regulator aven
NM_006311	NCOR1	Nuclear receptor co-repressor 1

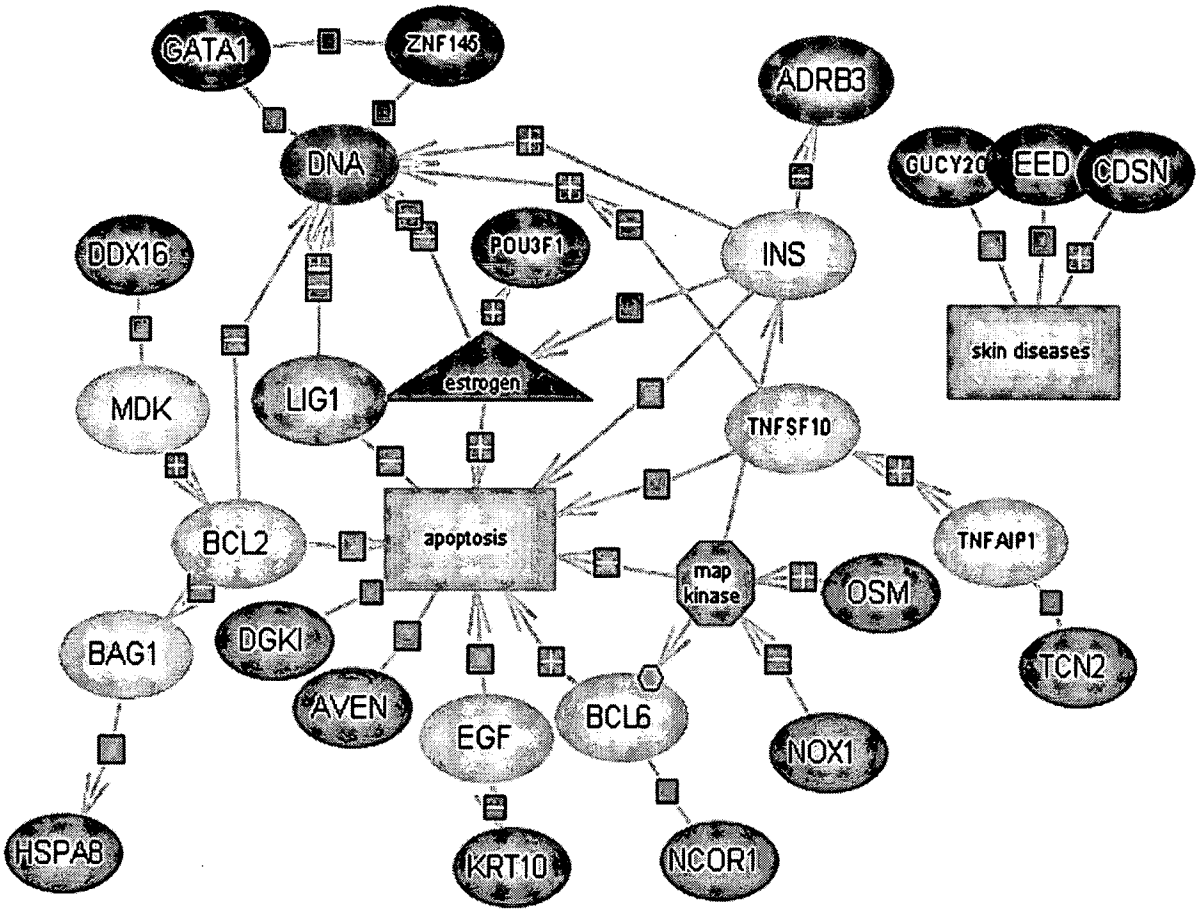


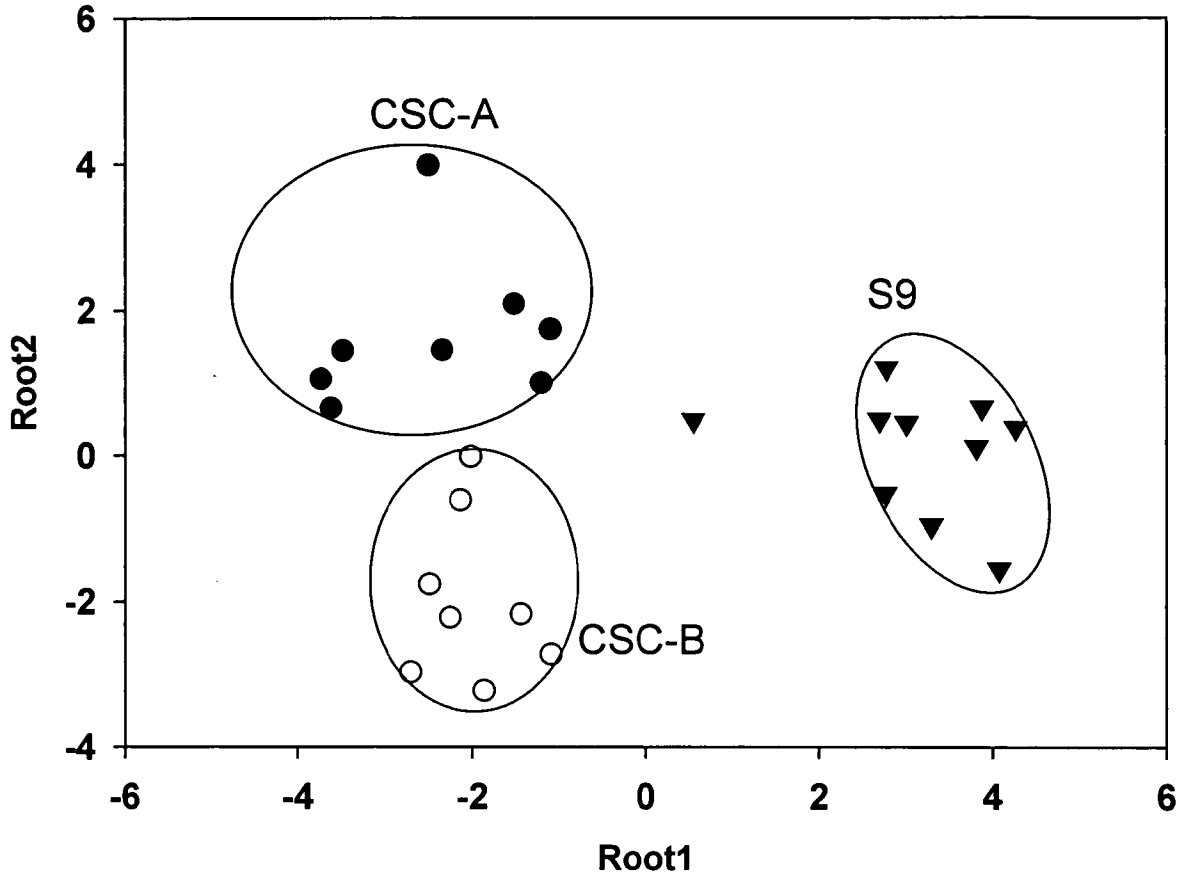
Figure 6. Functional associations of genes highly correlated in all three treatment groups. The genes, pathways, and functional interconnections among these elements for genes correlated in all three treatment groups are represented. Gene and pathway symbols are described in figure 5. Red ovals indicate genes from Table 3. Grey ovals (indicating additional proteins not in Table 3), yellow

oval (cell object – DNA) and green triangle (indicating small molecule - estrogen) were added to better define the regulatory networks of the genes identified in this analysis. Orange ovals indicate classes of functional peptides. Yellow rectangles indicate cellular processes in which these genes participate. Each line indicates a regulatory relationship (binding, regulation, etc.) based upon a literature reference. Regulatory relationships are denoted in a box on the line with positive regulation represented as a plus sign, negative regulation as a minus sign, and unknown relationships by no sign.

Refined Analysis of CSC-correlated Genes using Discriminant Function Analysis (DFA). The set of 40 genes that are correlated after CSC treatments (see Table 2 and Figure 4) but not correlated after S9 treatment were further analyzed using DFA. DFA is a form of multivariate analysis that identifies subsets of dependent variables that characterize a system made up of related groups. In this kind of expression analysis a linear equation is calculated, denoted a root, whose overall value is distinct for a given characterized group. DFA identifies genes most characteristic of a given state. Of the 40 CSC-correlated genes, 11 were identified by DFA as being most highly distinct among CSC and S9 treated cells (Table 4). Interestingly, a significant number of these genes are associated with oncogenesis. For example, this gene set includes 3 putative proto-oncogenes including 1) MAP2K5 the overexpression of which is associated with increased proliferative and invasive potential of metastatic prostate cancer and is reported to be a potent survival molecule in APO- MCF-7 breast carcinoma cells (58; 88); 2) DDIT3 a C/EBP transcriptional regulator involved in growth arrest induced by DNA damage that is a common breakpoint in human myxoid liposarcomas (17); and 3) BAG2 a BCL-2-binding apoptosis

suppressor that is overexpressed in human cervical, breast and lung cancer cell lines (93). In addition, three putative tumor suppressor genes were also identified in this gene set. These were FUS1, RASA1, and FPH2L1. FUS1 can inhibit tumor cell growth by inducing apoptosis (46), and was first identified in a search for potential tumor suppressors within a critical homozygous deletion region at 3p21.3 common in lung cancers (53). RASA1 as a key member of the GAP1 family of GTPase-activating proteins plays a key role in the Ras signaling pathway (5). DPH2L1 is a BRCA1-induced gene that maps within a region of 17p13.3, which is deleted in 80% of all ovarian epithelial malignancies. DPH2L1 was identified by exon trapping in this region and was implicated as a tumor suppressor as its expression is reduced or undetectable in ovarian tumors and tumor cell lines (3; 72; 77). In addition, a nicotinic cholinergic receptor, CHRNA9, and two putative neural growth factors, NxpH3, a neuropeptide-like neural signaling molecule (60), and NINJ2, a gene upregulated in damaged nerve cells that upregulates neurite outgrowth (2), were also identified in this gene set. The impact on neural growth factors is not surprising in light of the fact that many lung cancers express neuroendocrine features and are also stimulated by an autocrine/paracrine system of neuroendocrine peptide hormones (8; 10).

A graphical representation of the DFA results for the three treatment conditions at all time points was generated. The spatial organization of the elements in this representation provides a measure of the overall variance among groups (Fig. 7). The genes used for this analysis were correlated in CSCs and not correlated in S9. A correlation coefficient of 0.8 was used as a threshold for defining similarity. The expression of these genes should therefore be similar in CSC-treated cells. Indeed the two CSC groups are more closely associated than either CSC group is to S9. Of note, the samples from the CSC groups do not overlap, suggesting that the two CSC treatments elicit somewhat distinct responses even in genes highly correlated in their behavior in each CSC group.



s.

Figure 7. Discriminant function analysis (DFA) identified genes having high discriminatory capabilities. Values of the roots obtained by DFA analysis were used to graphically depict the differences of the gene expression values obtained for the three treatments (CSC-A, CSC-B, and S9). Root values for the 2-12h time points for each treatment are represented by filled circles (CSC-A), open circles (CSC-B), and filled triangles (S9).

Table 4: Discriminant Function Analysis of CSC-Correlated Genes

GenBank	Gene	Gene description

accession no.	abbreviation	
M23326	TRDV3	T cell receptor delta variable 3
NM_002757	MAP2K5	Mitogen-activated protein kinase kinase 5
NM_004083	DDIT3	DNA-damage-inducible transcript 3
NM_004282	BAG2	BCL2-associated athanogene 2
NM_007275	FUS1	Lung cancer candidate
NM_003408	ZFP37	Zinc finger protein 37 homolog (mouse)
NM_002046	GAPD	Glyceraldehyde-3-phosphate dehydrogenase
NM_017581	CHRNA9	Cholinergic receptor, nicotinic
BC015737	NINJ2	Ninjurin 2
AB032985	NXPH3	Neurexophilin 3
NM_002890	RASA1	RAS p21 protein activator
NM_001383	DPH2L1	Diptheria toxin resistance protein

Figure 8 shows the result of functional analysis of the gene set in Table 3.using Pathway Assist. Not surprisingly, the major cellular processes affected by these genes are a subset of the processes affected by the parent gene set, as illustrated in Figure 5.

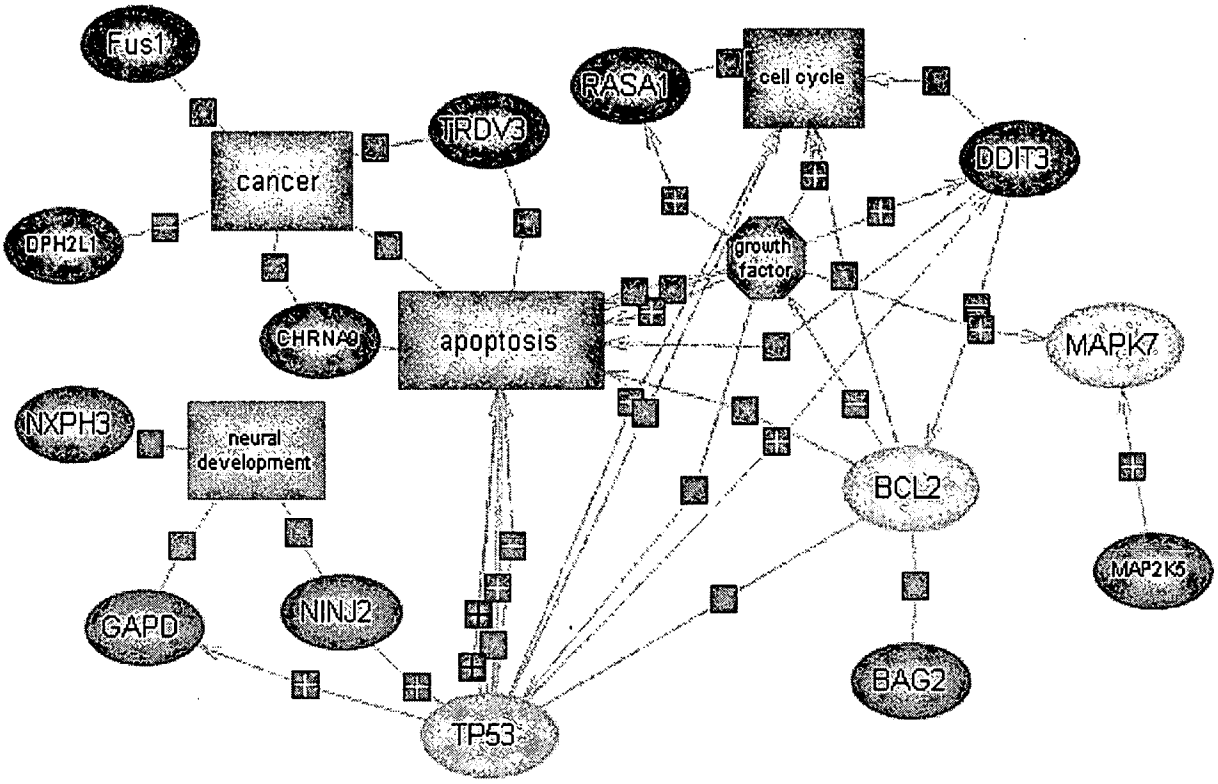


Figure 8. Functional associations of genes presented in Table 3. The genes, pathways, and functional interconnections among these elements for genes having the highest discriminatory potential among all three treatment groups are represented. Gene and pathway symbols are described in figure 5.

DISCUSSION

Relatively little is known about the effects of CSC exposure on overall impact on steady state mRNA levels and transcriptional regulation in normal lung cells. The fact that cigarette smoke, as well as various smoke components, can cause numerous disruptions to the genome (16; 90), transcriptome (6; 29), and proteome (35) presents the possibility of identifying a set of relevant biomarkers that could be useful for monitoring exposure to tobacco toxins, detecting premalignant disease, improving diagnosis and prognosis of current disease, developing new treatment options, and testing risk reduction strategies for current and former smokers (83; 92). In addition, elucidating the various molecular, genetic, cellular, and systemic effects of cigarette smoke should result in a detailed mechanistic understanding of how chronic tobacco exposure ultimately causes disease. Several studies assessing the clinical usefulness of alterations in global gene and protein expression patterns in malignant and normal human lung tissues have shown that quantitative and/or qualitative changes in a small number of expressed genes and proteins, in combination with standard clinicopathological variables, may have prognostic and/or diagnostic potential in patients with tobacco-related diseases (6; 14; 29; 32; 35; 57; 62; 66; 94; 98). However, a direct cause and effect relationship between any of these documented molecular events and cell exposure to tobacco smoke is unclear (76; 92). Thus, one relevant strategy is to examine the effects of tobacco constituents on the transcriptome of normal lung cells in a controlled in vitro environment. Using high-density microarrays and a novel method for analyzing array data, we show here that exposing human bronchial epithelial cells to cigarette smoke condensates from two commercial brands of American cigarettes identifies a set of genes whose expression levels vary over the normal variability of gene expression in these cells, and may therefore be indicators of tobacco-induced changes. Further, by sorting these genes into biologically functional classes, dominant biochemical pathways known to be relevant to tobacco-

related diseases were found to be responsive to tobacco condensate exposure. Finally, we note the surprising finding that treatment with an S9 fraction of metabolic enzymes, a step common to many toxicological studies, has a broad impact on gene expression in normal lung cells that is distinctly different from the effects of tobacco condensates. The novel data mining technique, termed hypervariable (HV) analysis, that we employed in this study has several strengths over conventional correlation and cluster assessment of microarray data, the most important being that monitoring hypervariability more accurately reflects subtle alterations to global gene expression patterns common to multiple conditions and allows them to be compared. This approach is particularly important since it appears that the expression level of most genes characteristically are altered less than 20% from baseline conditions as a result of physiologically relevant manipulations (38).

We established four post-treatment expression characteristics for each gene on the array: 1) whether or not the gene was expressed above background at each time-point; 2) whether or not the gene showed hypervariability (i.e. change greater than normal) of expression in one, two, or all three treatment conditions over the 12h treatment period; 3) what was the specific pattern of gene expression over the 12 h treatment period; and 4) whether or not the gene expression pattern in each condition correlated with its behavior under the two other conditions from 0-12h. Several interesting observations emerged from this study, the most important being that treatment of NHBE cells with CSCs from two American brands of cigarettes altered the expression of approximately 3600 genes and ORFs (or 17% of the array) sometime during the 12-hour exposure (see Figures 1 and 2). These data support our conjecture that due to their chemical complexity and temporal requirement for metabolic activation, CSCs should have a broad and dynamic effect on the homeostatic transcriptome of the NHBE cell. In addition to the quantitative similarities in gene alterations induced by the different CSCs, there are also

qualitative similarities in that both CSCs affect a large common block of genes, which is not surprising given the relatively comparable types of blended tobaccos used in most American cigarette brands. Clearly, deciphering the specific biological effect of each of the genes impacted by CSCs is desirable but not practical. However, with the use of several types of approaches to discriminate and cluster genes that became hypervariable after CSC treatment, it is possible to provide relatively robust and accurate statistical estimations of functional significance for these sets of variable genes, which can then be used to build biologically and clinically relevant models that can be tested. For example, as shown in Figure 5, CSCs affect networks of genes that intersect critical signaling pathways such as apoptosis, transcription, and cell cycle regulation, which are known to play key roles in specific diseases such as cancer, chronic inflammation, and impaired neural development, and which both epidemiological and functional studies conclude can be caused by chronic cigarette smoking. The relevance of these pathways to smoking-related diseases is further supported by a limited body of published data in which other cell types or tissues exposed to either smoke, CSC, or a specific substance in CSC (e.g., benzo[a]pyrene, nicotine, etc.) were assessed using low-density arrays (4; 30; 63; 65; 96).

The sensitivity and accuracy of the methodologies used in this study to identify genes impacted by CSCs is further shown by the fact that the set of HV genes in CSC-treated cells includes many of the genes and/or gene families that have been previously discovered using various global expression analyses (e.g., Serial Analysis of Gene Expression, Differential Display, and microarrays) and concluded to be of importance in the development and/or maintenance of lung cancers. These include erb-B2 (18), matrix metalloproteinase 9 (MMP9) (73), the heterogeneous nuclear ribonucleoprotein (hnRNP) family (79; 81; 84; 97), the Fus1 lung cancer candidate (46; 50; 93), glutathione S-transferase pi (85), the β-retinoic acid receptor (54; 87), chromogranin B (70), RAB5 (95), death-associated protein kinase 1 (DAPK) (36),

various cancer/testis antigens [MAGE genes (82)], and others. For many of these genes, however, the current study is the first to show that their expression is altered in normal bronchial epithelial cells exposed to CSCs for only a short period of time, which may indicate the one or more of these genes may be an early indicator of tobacco-related cellular damage. In addition, our data also detail a large number of genes and gene families that have not yet been identified as being relevant to the induction or maintenance of pulmonary neoplasms or to other tobacco-related diseases involving the cardiovascular and immune systems. One or more of these genes may prove to be novel biomarkers in the pathogenesis of these diseases. Therefore, these array data present the most detailed account of molecular effects of CSCs to date and may be instrumental in developing a new generation of candidate target genes for which functional models of cigarette smoke-affected biological pathways, gene interactions, and clinical relationships can be constructed and tested. This possibility is particularly important since, as yet, there is no single lung cancer biomarker that has achieved sufficient diagnostic significance to be of primary use in the clinic (42).

The highly correlated expression characteristics of the CSC-impacted genes shown in Table 1 and Figure 5 highlight a set of genes of which several appear to play prominent roles in tobacco-related diseases. For example, both DPH2L1 (13) and Fus1 (46) are putative tumor suppressor genes associated with ovarian and lung cancer, respectively. Fus1 is found at a homozygous deleted region of chromosome 3p21 in lung tumors (46), and its forced expression in lung carcinoma cells suppresses cell growth in vitro and growth and metastases of tumors in vivo by mechanisms involving G1-arrest and induction of apoptosis (46; 49). The RASA1 is a component of the GAP1 family of GTPase-activating proteins, which can suppress proliferation signals by enhancing the weak intrinsic GTPase activity of normal RAS p21 protein and maintaining it in its inactive GDP-bound form (68). Evidence suggests that Ras acts as a major

nexus for multiple signaling pathways that control a diverse range of functions (68), but many of the subtleties of Ras functioning in individual cell types remain unclear. Recent data suggest that it may have an important role in tumor cell survival (52). The MAP2K5 is a novel mitogen activated protein kinase implicated in the regulation of cell proliferation (20). Overexpression of MAP2K5 can, in cooperation with other effectors, transform rodent cells (71), and function as a potent survival molecule in breast cancer cells (89). MAP2K5 represents a potential therapeutic target in prostate cancer as overexpression of MAP2K5 can induce proliferation, motility, and invasion (59). Interestingly, MAP2K5 also dramatically upregulates the expression of matrix metalloproteinase-9 (MMP-9) in prostate cancers (59). As shown in Table 1, we found that MMP-9 is HV in both CSC-treatment groups. The matrix metalloproteinases (MMPs) are a large family of extracellular matrix degrading enzymes believed to play central roles in degradation, remodeling, and repair of basement membranes. Inappropriate or overexpression of these proteins appear to a critical determinant in tumor invasion and metastasis of a number of neoplasms including those of the lung (67). For example, MMP9 potentiates pulmonary metastasis formation (41; 86), and high serum levels of MMP-9 in patients with non-small-cell lung cancer (NSCLC) correlated with significantly shorter survival than patients with low serum levels of this protein (51). Based on these and other studies, the assessment of drugs that inhibit MMP-9 as an adjuvant approach is the focus of several clinical trials in patients with lung cancers (25).

In addition to a common set of affected genes, each CSC also altered the expression of a relatively large gene set that was unique to each CSC. The impact on these unique gene sets may be due to qualitative and/or quantitative differences in the constellation of chemical constituents in the two CSCs. We note that despite the fact that both Brand A and Brand B are similar types of cigarettes (i.e., 'full-flavor') as determined by FTC criteria, there are measurable differences

in the quantities of nicotine, tar, and a small panel of toxins and carcinogens between Brand-A and Brand-B cigarettes (unpublished data). Whether the differences in one or more of these substances directly correlates with the observed gene changes at the mRNA level remains to be determined. Moreover, whether the unique gene sets affected by CSC-A and CSC-B ultimately influence different cellular pathways and induce different biological phenomena also requires further research. Several basic assumptions of the emerging field of toxicogenomics is that there are reasonable similarities in gene expression patterns induced by multiple members of one specific class of toxicants, and subtle differences in these gene expression patterns may distinguish distinct chemical-specific 'gene signatures' of exposure (1; 64).

Figure 2, which shows the major temporal changes in gene expression, indicates that the majority of CSC-affected genes do not return to baseline within the 12-hour treatment period, especially for CSC-B-affected genes. This phenomenon could be due simply to the fact that the cells were chronically exposed to the CSCs for the entire 12-hours. However, a more biologically relevant possibility is that many of the affected genes may require a significant amount of time to return to baseline even after exposure is terminated. If this scenario is valid in vivo, then the current pack-a-day smoker who averages >150 cigarette puffs/day may alter the homeostatic expression of a large number of genes that cannot return to a baseline state during a typical day. One speculation is that the chronically perturbed state (either increased or decreased compared to baseline) of one or more of these genes may ultimately be etiologically involved in various pathological states caused by exposure to cigarette smoke. Indirect support for this idea comes from the fact that in subjects who quit smoking there is both short-term improvement in the functioning of a number of affected organ systems (e.g., lung, cardiovascular structures, kidneys, etc.) and a long-term decline in incidence and mortality from various diseases affecting these systems (24). Presumably, this reversal of smoking-related damage at the tissue and

population levels reflects a corresponding reversal at a molecular and cellular level. For example, it is well documented that chronic inflammatory processes in smokers play fundamental roles in the pathogenesis of atherosclerosis, and increased plasma and tissue levels of several biomarkers associated with inflammation such as various cytokines (e.g., IL-1β, TNF-α), pro-atherogenic enzymes (e.g., lipoprotein lipase) and cell adhesion molecules (e.g., VCAM-1) are associated with future cardiovascular risk (7; 27), while smoking cessation leads to decreased expression of many pro-inflammatory biomolecules and a concomitant reduction in cardiovascular risk (7; 27). It is also possible that the altered expression of one or more genes in the habitual smoker becomes attenuated with time as an adaptive response to the stress of chronic activation, and this phenomenon may have unanticipated long-term biological consequences for the smoker (30). Clearly, understanding the complex toxicogenetic relationships between chronic long-term exposure to a mixture of tobacco carcinogens/toxins and temporal patterns of molecular and cellular dysfunctions will yield important insights into disease mechanisms and lead to the identification of reliable risk-related biomarkers.

One unexpected finding of this study was the relatively broad effect of the S9 metabolic enzyme fraction on gene expression in NHBE cells. S9-exposed cells are traditionally considered a negative control for toxicogenetic experiments performed to establish environmental and occupational exposure guidelines (28). The fact that we observed gene alterations as early as 2 hours post-S9 exposure has interpretive implications for standard toxicological assays that routinely measure biological and genetic effects of control and test substances after 4 hours of exposure. This observation is particularly relevant as the global shift towards advanced genomic and proteomic technologies transforms the field of toxicology from one relying on the induction of gross genetic abnormalities such as mutations and structural/numerical chromosomal abnormalities to one where altered expression of panels of

genes and proteins are used to determine risk to the human population. In order to clearly establish the potential toxicity or efficacy of an environmental substance, drug, or chemopreventive agent, it will be important to prove that control substances or vehicles cause minimal disruption of the physiologically normal transcriptome. Furthermore, since S9 can induce a range of alterations in gene expression levels independent of any test substance, it is possible that one or more S9-induced effects can be synergistic or antagonistic with the test substances (19; 78). For example, Figure 3 shows that many of the same genes that are down-regulated in S9-treated cells are upregulated in CSC-treated cells despite the fact that CSCs contain the same concentration of S9 enzymes. Alternatively, the effects of S9 can be mitigated by the test substance. Evidence for this possibility is strongly supported by our data, which shows that a number of genes whose steady-state mRNA level were found to be altered only by S9 were not found to be altered when cells were exposed to S9 in context with either CSC-A or CSC-B. In this scenario, the direct effects of S9, which can be directly cytotoxic to cells in cultures (48), may be attenuated when sequestered and modified through contact with substances in CSCs.

The current study attempted to define, in broad outline, the range of perturbations to the homeostatic transcriptome of the NHBE cells as a prelude to understanding pathogenetic events occurring in chronic smokers. It is a reasonable hypothesis that one or more of the genes whose expression status are altered by exposure to CSC, and/or one or more of the biological pathways in which these genes function, is permanently disabled or permuted in vivo in smokers, which may contribute to specific steps in the pathogenesis of a tobacco-related disease. Thus, the data presented in this paper provide a working atlas of tobacco-specific effects on normal lung cells and suggests various molecular routes that can be assessed in future studies for direct or indirect roles in disease.

References

1. **Afshari CA, Nuwaysir EF and Barrett JC**. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res* 59: 4759-4760, 1999.

2. **Araki T and Milbrandt J**. Ninjurin2, a novel homophilic adhesion molecule, is expressed in mature sensory and enteric neurons and promotes neurite outgrowth. *J Neurosci* 20: 187-195, 2000.

3. **Atalay A, Crook T, Ozturk M and Yulug IG**. Identification of genes induced by BRCA1 in breast cancer cells. *Biochem Biophys Res Commun* 299: 839-846, 2002.

4. **Bartosiewicz M, Penn S and Buckpitt A**. Applications of gene arrays in environmental toxicology: fingerprints of gene regulation associated with cadmium chloride, benzo(a)pyrene, and trichloroethylene. *Environ Health Perspect* 109: 71-74, 2001.

5. **Bernards A**. GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and Drosophila. *Biochim Biophys Acta* 1603: 47-82, 2003.

6. **Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ and Meyerson M**. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 98: 13790-13795, 2001.

7. **Blake GJ and Ridker PM**. Inflammatory bio-markers and cardiovascular risk prediction. *J Intern Med* 252: 282-294, 2002.

8. **Brambilla E, Travis WD, Colby TV, Corrin B and Shimosato Y**. The new World Health Organization classification of lung tumours. *Eur Respir J* 18: 1059-1068, 2001.

9. **Bunn PA, Jr., Helfrich BA, Brenner DG, Chan DC, Dykes DJ, Cohen AJ and Miller YE**. Effects of recombinant neutral endopeptidase (EC 3.4.24.11) on the growth of lung cancer cell lines in vitro and in vivo. *Clin Cancer Res* 4: 2849-2858, 1998.

10. **Carnaghi C, Rimassa L, Garassino I and Santoro A**. Clinical significance of neuroendocrine phenotype in non-small-cell lung cancer. *Ann Oncol* 12 Suppl 2: S119-S123, 2001.

11. **Chang ML, Eddy RL, Shows TB and Lau JT**. Three genes that encode human beta-galactoside alpha 2,3-sialyltransferases. Structural analysis and chromosomal mapping studies. *Glycobiology* 5: 319-325, 1995.

12. **Chau BN, Cheng EH, Kerr DA and Hardwick JM**. Aven, a novel inhibitor of caspase activation, binds Bcl-xL and Apaf-1. *Mol Cell* 6: 31-40, 2000.

13. **Chen CM and Behringer RR**. Cloning, structure, and expression of the mouse Ovca1 gene. *Biochem Biophys Res Commun* 286: 1019-1026, 2001.

14. **Chen G, Gharib TG, Huang C-C, Thomas DG, Shedden KA, Taylor JMG, Kardia SLR, Misek DE, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM and Beer DG**. Proteomic analysis of lung adenocarcinomas: identification of a highly expressed set of proteins in tumors. *Clin Cancer Res* 8: 2298-2305, 2002.

15. **Cho JH, Lee D, Park JH, Kim K and Lee IB**. Optimal approach for classification of acute leukemia subtypes based on gene expression data. *Biotechnol Prog* 18: 847-854, 2002.

16. **Chujo M, Noguchi T, Miura T, Arinaga M, Uchida Y and Tagawa Y**. Comparative genomic hybridization analysis detected frequent overrepresentation of chromosome 3q in squamous cell carcinoma of the lung. *Lung Cancer* 38: 23-29, 2002.

17. **Crozat A, Aman P, Mandahl N and Ron D**. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* 363: 640-644, 1993.

18. **D'Amico TA, Massey M, Herndon JE2, Moore MB and Harpole DH, Jr.** A biologic risk model for stage I lung cancer: immunohistochemical analysis of 408 patients with the use of ten molecular markers. *J Thorac Cardiovasc Surg* 117: 736-743, 1999.

19. **Dertinger SD, Torous DK and Tometsko AM**. In vitro system for detecting non-genotoxic carcinogens. *Environ Mol Mutagen* 21: 332-338, 1993.

20. **Diaz-Meco MT and Moscat J**. MEK5, a new target of the atypical protein kinase C isoforms in mitogenic signaling. *Mol Cell Biol* 21: 1218-1227, 2001.

21. **Dozmorov I and Centola M.** An associative analysis of gene expression array data. *Bioinformatics* 19: 204-211, 2003.

22. **Dozmorov I, Galecki A, Chang Y, Krzesicki R, Vergara M and Miller RA.** Gene expression profile of long-lived snell dwarf mice. *J Gerontol A Biol Sci Med Sci* 57: B99-108, 2002.

23. **Dozmorov I, Saban MR, Gerard NP, Lu B, Nguyen NB, Centola M and Saban R.** Neurokinin 1 receptors and neprilysin modulation of mouse bladder gene regulation. *Physiol Genomics* 12: 239-250, 2003.

24. **Fagerstrom K.** The epidemiology of smoking: health consequences and benefits of cessation. *Drugs* 62: 1-9, 2002.

25. **Falardeau P, Champagne P, Poyet P, Hariton C and Dupont E.** Neovastat, a naturally occurring multifunctional antiangiogenic drug, in phase III clinical trials. *Semin Oncol* 28: 620-625, 2001.

26. Federal Register. Volume 32, Number 147, 11-178. 8-1-1967.

Ref Type: Report

27. **Frohlich M, Sund M, Lowel H, Imhof A, Hoffmeister A and Koenig W.** Independent associatons of various smoking characteristics with markers of systemic inflammation in men. Results from a representative sample of the general population (MONICA Augsburg survey 1994/95). *Eur Heart J* 24: 1365-1372, 2003.

28. **Galloway SM, Aardema MJ, Ishidate M, Jr., Ivett JL, Kirkland DJ, Morita T, Mosesso P and Sofuni T**. Report from working group on in vitro tests for chromosomal aberrations. *Mutat Res* 312: 241-261, 1994.

29. **Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de RM, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D and Petersen I**. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 98: 13784-13789, 2001.

30. **Gebel S, Gerstmayer B, Bosio A, Haussmann HJ, Van Miert E and Muller T**. Gene expression profiling in respiratory tissues from rats exposed to mainstream cigarette smoke. *Carcinogenesis* 2003.

31. **Gibbs S and Ponec M**. Intrinsic regulation of differentiation markers in human epidermis, hard palate and buccal mucosa. *Arch Oral Biol* 45: 149-158, 2000.

32. **Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardia SL, Beer DG, Rennert G, Cho KR, Gruber SB, Fearon ER and Hanash S**. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 159: 1231-1238, 2001.

33. **Glynne RJ, Ghandour G and Goodnow CC**. Genomic-scale gene expression analysis of lymphocyte growth, tolerance and malignancy. *Curr Opin Immunol* 12: 210-214, 2000.

34. **Guengerich FP**. Metabolism of chemical carcinogens. *Carcinogenesis* 21: 345-351, 2000.

35. **Hanash S, Brichory F and Beer D**. A proteomic approach to the identification of lung cancer markers. *Dis Markers* 17: 295-300, 2001.

36. **Harden SV, Tokumaru Y, Westra WH, Goodman S, Ahrendt SA, Yang SC and Sidransky D**. Gene promoter hypermethylation in tumors and lymph nodes of stage I lung cancer patients. *Clin Cancer Res* 9: 1370-1375, 2003.

37. **Harrington JJ, Sherf B, Rundlett S, Jackson PD, Perry R, Cain S, Leventhal C, Thornton M, Ramachandran R, Whittington J, Lerner L, Costanzo D, McElligott K, Boozer S, Mays R, Smith E, Veloso N, Klika A, Hess J, Cothren K, Lo K, Offenbacher J, Danzig J and Ducar M**. Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat Biotechnol* 19: 440-445, 2001.

38. **Hatfield GW, Hung SP and Baldi P**. Differential analysis of DNA microarray gene expression data. *Mol Microbiol* **47**: 871-877, 2003.

39. **Hecht SS**. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol* 3: 461-469, 2002.

40. **Hecht SS.** Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature Reviews: Cancer* 3: 733-744, 2003.

41. **Hiratsuka S, Nakamura K, Iwai S, Murakami M, Itoh T, Kijima H, Shipley JM, Senior RM and Shibuya M.** MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell* 2: 289-300, 2002.

42. **Hirsch FR, Franklin WA, Gazdar AF and Bunn PA, Jr.** Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. *Clin Cancer Res* 7: 5-22, 2001.

43. **Hoffmann D and Hecht SS.** Advances in tobacco carcinogenesis. In: Handbook of Experimental Pharmacology, edited by Cooper CS and Grover PL. Heidelberg: Springer-Verlag, 1990, p. 63-102.

44. **Inatani M and Yamaguchi Y.** Gene expression of EXT1 and EXT2 during mouse brain development. *Brain Res Dev Brain Res* 141: 129-136, 2003.

45. **International Agency for Research on Cancer.** *Tobacco Smoking. IARC Monographs on the Evaulation of the Carcinogenic Risk of Chemicals to Humans.* IARC, Lyon: 1986.

46. **Ji L, Nishizaki M, Gao B, Burbee D, Kondo M, Kamibayashi C, Xu K, Yen N, Atkinson EN, Fang B, Lerman MI, Roth JA and Minna JD.** Expression of several genes in the human chromosome 3p21.3 homozygous deletion region by an adenovirus

vector results in tumor suppressor activities in vitro and in vivo. *Cancer Res* 62: 2715-2720, 2002.

47. **Kerr MK, Martin M and Churchill GA**. Analysis of variance for gene expression microarray data. *J Comput Biol* 7: 819-837, 2000.

48. **Kohn J and Durham HD**. S9 liver fraction is cytotoxic to neurons in dissociated culture. *Neurotoxicology* 14: 381-386, 1993.

49. **Kondo M, Ji L, Kamibayashi C, Tomizawa Y, Randle D, Sekido Y, Yokota J, Kashuba V, Zabarovsky E, Kuzmin I, Lerman M, Roth J and Minna JD**. Overexpression of candidate tumor suppressor gene FUS1 isolated from the 3p21.3 homozygous deletion region leads to G1 arrest and growth inhibition of lung cancer cells. *Oncogene* 20: 6258-6262, 2001.

50. **Kondo M, Ji L, Kamibayashi C, Tomizawa Y, Randle D, Sekido Y, Yokota J, Kashuba V, Zabarovsky EK, Uzmin I, Lerman M, Roth J and Minna JD**. Overexpression of candidate tumor suppressor gene FUS1 isolated from the 3p21.3 homozygous deletion region leads to G1 arrest and growth inhibition of lung cancer cells. *Oncogene* 20: 6258-6262, 2001.

51. **Laack E, Kohler A, Kugler C, Dierlamm T, Knuffmann C, Vohwinkel G, Niestroy A, Dahlmann N, Peters A, Berger J, Fiedler W and Hossfeld DK**. Pretreatment serum

levels of matrix metalloproteinase-9 and vascular endothelial growth factor in non-small-cell lung cancer. *Ann Oncol* 13: 1550-1557, 2002.

52. **Leblanc V, Delumeau I and Tocque B**. Ras-GTPase activating protein inhibition specifically induces apoptosis of tumour cells. *Oncogene* 18: 4884-4889, 1999.

53. **Lerman MI and Minna JD**. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. *Cancer Res* 60: 6116-6133, 2000.

54. **Lotan R**. Roles of retinoids and their nuclear receptors in the development and prevention of upper aerodigestive tract cancers. *Environ Health Perspect* 105: 985-988, 1997.

55. **Lustig LR and Peng H**. Chromosome location and characterization of the human nicotinic acetylcholine receptor subunit alpha (alpha) 9 (CHRNA9) gene. *Cytogenet Genome Res* 98: 154-159, 2002.

56. **Lyda MH and Weiss LM**. Immunoreactivity for epithelial and neuroendocrine antibodies are useful in the differential diagnosis of lung carcinomas. *Hum Pathol* 31: 980-987, 2000.

57. **McDoniels-Silvers AL, Stoner GD, Lubet RA and You M.** Differential expression of critical cellular genes in human lung adenocarcinomas and squamous cell carcinomas in comparison to normal lung tissues. *Neoplasia* 4: 141-150, 2002.

58. **Mehta PB, Jenkins BL, McCarthy L, Thilak L, Robson CN, Neal DE and Leung HY.** MEK5 overexpression is associated with metastatic prostate cancer, and stimulates proliferation, MMP-9 expression and invasion. *Oncogene* 22: 1381-1389, 2003.

59. **Mehta PB, Jenkins BL, McCarthy L, Thilak L, Robson CN, Neal DE and Leung HY.** MEK5 overexpression is associated with metastatic prostate cancer, and stimulates proliferation, MMP-9 expression and invasion. *Oncogene* 22: 1381-1389, 2003.

60. **Missler M and Sudhof TC.** Neurexophilins form a conserved family of neuropeptide-like glycoproteins. *J Neurosci* 18: 3630-3638, 1998.

61. **Moore JH, Parker JS, Olsen NJ and Aune TM.** Symbolic discriminant analysis of microarray data in autoimmune disease. *Genet Epidemiol* 23: 57-69, 2002.

62. **Nacht M, Dracheva T, Gao Y, Fujii T, Chen Y, Player A, Akmaev V, Cook B, Dufault M, Zhang M, Zhang W, Guo M, Curran J, Han S, Sidransky D, Buetow K, Madden SL and Jen J.** Molecular characteristics of non-small cell lung cancer. *Proc Natl Acad Sci U S A* 98: 15203-15208, 2001.

63. **Nadadur SS, Pinkerton KE and Kodavanti UP**. Pulmonary gene expression profiles of spontaneously hypertensive rats exposed to environmental tobacco smoke. *Chest* 121: 83S-84S, 2002.

64. **Neumann NF and Galvez F**. DNA microarrays and toxicogenomics: applications for ecotoxicology. *Biotechnol Adv* 20: 391-419, 2002.

65. **Nordskog BK, Blixt AD, Morgan WT, Fields WR and Hellmann GM**. Matrix-degrading and pro-inflammatory changes in human vascular endothelial cells exposed to cigarette smoke condensate. *Cardiovasc Toxicol* 3: 101-117, 2003.

66. **Oh JM, Brichory F, Puravs E, Kuick R, Wood C, Rouillard JM, Tra J, Kardia S, Beer D and Hanash S**. A database of protein expression in lung cancer. *Proteomics* 1: 1303-1319, 2001.

67. **Ohbayashi H**. Matrix metalloproteinases in lung diseases. *Curr Protein Pept Sci* 3: 409-421, 2002.

68. **Olson MF and Marais R**. Ras protein signalling. *Semin Immunol* 12: 63-73, 2000.

69. **Opalka B, Dickopp A and Kirch HC**. Apoptotic genes in cancer therapy. *Cells Tissues Organs* 172: 126-132, 2002.

70. **Parriarca C, Pruneri G, Alfano RM, Carboni N, Ermellino L, Guddo F, Buffa R, Siccardi AG and Coggi G**. Polysialylated N-CAM, chromogranin A and B, and

secretogranin II in neuroendocrine tumors of the lung. *Virchows Arch* 430: 455-460, 1997.

71. **Pearson G, English JM, White MA and Cobb MH**. ERK5 and ERK2 cooperate to regulate NF-kappaB and cell transformation. *J Biol Chem* 276: 7927-7931, 2001.

72. **Phillips NJ, Zeigler MR and Deaven LL**. A cDNA from the ovarian cancer critical region of deletion on chromosome 17p13.3. *Cancer Lett* 102: 85-90, 1996.

73. **Pritchard SC, Nicolson MC, Lloret C, McKay JA, Ross VG, Kerr KM, Murray GI and McLeod HL**. Expression of matrix metalloproteinases 1,2, 9 and their tissue inhibitors in stage II non-small cell lung cancer: implications for MMP inhibition therapy. *Oncol Rep* 8: 421-424, 2001.

74. **Rocke DM and Durbin B**. A model for measurement error for gene expression arrays. *J Comput Biol* 8: 557-569, 2001.

75. **Rodgman A, Smith CJ and Perfetti TA**. The composition of cigarette smoke: a retrospective, with emphasis on polycyclic components. *Hum Exp Toxicol* 19: 573-595, 2000.

76. **Rom WN and Tchou-Wong KM**. Functional genomics in lung cancer and biomarker detection. *Am J Respir Cell Mol Biol* 29: 157-162, 2003.

77. **Schultz DC, Vanderveer L, Berman DB, Hamilton TC, Wong AJ and Godwin AK.** Identification of two candidate tumor suppressor genes on chromosome 17p13.3. *Cancer Res* 56: 1997-2002, 1996.

78. **Shah AB, Combes RD and Rowland IR.** Interaction with microsomal lipid as a major factor responsible for S9-mediated inhibition of 1,8-dinitropyrene mutagenicity. *Mutat Res* 249: 93-104, 1991.

79. **Snead DR, Perunovic B, Cullen N, Needham M, Dhillon DP, Satoh H and Kamma H.** hnRNP B1 expression in benign and malignant lung disease. *J Pathol* 200: 88-94, 2003.

80. **Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D and Futcher B.** Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* 9: 3273-3297, 1998.

81. **Sueoka E, Goto Y, Sueoka N, Kai Y, Kozu T and Fujiki H.** Heterogeneous nuclear ribonucleoprotein B1 as a new marker of early detection for human lung cancers. *Cancer Res* 59: 1404-1407, 1999.

82. **Tajima K, Obata Y, Tamaki H, Yoshida M, Chen YT, Scanlan MJ, Old LJ, Kuwano H, Takahashi T, Takahashi T and Mitsudomi T.** Expression of cancer/testis (CT) antigens in lung cancer. *Lung Cancer* 42: 23-33, 2003.

83. **Thun MJ, Henley SJ and Calle EE**. Tobacco use and cancer: an epidemiologic perspective for geneticists. *Oncogene* 21: 7307-7325, 2002.

84. **Tockman MS, Mulshine JL, Piantadosi S, Erozan YS, Gupta PK, Ruckdeschel JC, Taylor PR, Zhukov T, Zhou WH, Qiao YL and Yao SX**. Prospective detection of preclinical lung cancer: results from two studies of heterogenous nuclear ribonucleoprotein A2/B1 overexpression. *Clin Cancer Res* 3: 2237-2246, 1997.

85. **Toyooka SS, Toyooka KO, Maruyama R, Virmani AK, Girard L, Miyajima K, Harada K, Ariyoshi Y, Takahashi T, Sugio K, Brambilla E, Gilcrease M, Minna JD and Gazdar AF**. DNA methylation profiles of lung tumors. *Mol Cancer Ther* 103: 61-67, 2001.

86. **van Kempen LC and Coussens LM**. MMP9 potentiates pulmonary metastasis formation. *Cancer Cell* 2: 251-252, 2002.

87. **Virmani AK and Gazdar AF**. Tumor suppressor genes in lung cancer. *Methods Mol Biol* 222: 97-115, 2003.

88. **Weldon CB, Scandurro AB, Rolfe KW, Clayton JL, Elliott S, Butler NN, Melnik LI, Alam J, McLachlan JA, Jaffe BM, Beckman BS and Burow ME**. Identification of mitogen-activated protein kinase kinase as a chemoresistant pathway in MCF-7 cells by using gene expression microarray. *Surgery* 132: 293-301, 2002.

89. **Weldon CB, Scandurro AB, Rolfe KW, Clayton JL, Elliott S, Butler NN, Melnik LI, Alam J, McLachlan JA, Jaffe BM, Beckman BS and Burow ME**. Identification of mitogen-activated protein kinase kinase as a chemoresistant pathway in MCF-7 cells by using gene expression microarray. *Surgery* 132: 293-301, 2002.

90. **Wistuba I, Gazdar AF and Minna JD**. Molecular genetics of small cell lung carcinoma. *Semin Oncol* 28: 3-13, 2001.

91. **Wynder EL and Hoffmann D**. Re: Cigarette smoking and the histopathology of lung cancer. *J Natl Cancer Inst* 90: 1486-1488, 1998.

92. **Yanagisawa K, Xu BJ, Carbone DP and Caprioli RM**. Molecular fingerprinting in human lung cancer. *Clin Lung Cancer* 5: 113-118, 2003.

93. **Yang X, Chernenko G, Hao Y, Ding Z, Pater MM, Pater A and Tang SC**. Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells. *Oncogene* 17: 981-989, 1998.

94. **Yoneda K, Peck K, Chang MM, Chmiel K, Sher YP, Chen J, Yang PC, Chen Y and Wu R**. Development of high-density DNA microarray membrane for profiling smoke- and hydrogen peroxide-induced genes in a human bronchial epithelial cell line. *Am J Respir Crit Care Med* 164: S85-S90, 2001.

95. **Yu L, Hui-chen F, Chen Y, Zou R, Yan S, Chun-xiang L, Wu-ru W and Li P.** Differential expression of RAB5A in human lung adenocarcinoma cells with different metastasis potential. *Clin Exp Metastasis* 17: 213-219, 1999.

96. **Zhang S, Day IN and Ye S.** Microarray analysis of nicotine-induced changes in gene expression in endothelial cells. *Physiol Genomics* 5: 187-192, 2001.

97. **Zhou J, Nong L, Wloch M, Cantor A, Mulshine JL and Tockman MS.** Expression of early lung cancer detection marker: hnRNP-A2/B1 and its relation to microsatellite alteration in non-small cell lung cancer. *Lung Cancer* 34: 341-350, 2001.

98. **Zhukov TA, Johanson RA, Cantor AB, Clark RA and Tockman MS.** Discovery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. *Lung Cancer* 40: 267-279, 2003.

ABSTRACT

Purpose: **This study assessed the impact on gene expression patterns in normal human bronchial epithelial (NHBE) cells exposed to cigarette smoke condensates (CSC) from commercial cigarettes. The ultimate goal of these studies is to develop a precise understanding of the genomic impact of tobacco smoke exposure, and to define biomarkers that can potentially discriminate tobacco-related effects and outcomes in a clinical setting.**

Experimental Design: **NHBE cells were treated with smoke condensates (200 ug/ml) from two American brands of cigarettes for up to 12 hours in the presence of S9 microsomal fraction from Aroclor 1254-treated rats. High-density oligonucleotide microarrays coupled with a novel statistical analysis that relies on statistical significance levels rather than arbitrary fold-change differences was used to identify genes that undergo qualitative and quantitative alterations in expression upon CSC treatment.**

Results: **A set of approximately 3700 genes was identified whose expression patterns altered over time after treatment with CSCs. While a majority of the genes in this set was affected by both condensates, each condensate also affected a unique subset of ~1000 genes. An unexpected finding was that the S9 microsomal fraction, required for metabolizing the procarcinogens in CSCs to carcinogenic metabolites, alone altered the expression of a large set of approximately 1700 genes, the majority of which overlapped with the tobacco-affected gene sets.**

Conclusions: **Exposure of NHBE cells to condensates from two brands of cigarettes alters the expression of a large common set of genes primarily in similar ways. Moreover, the data indicate that both condensates affect a common set of biological pathways, including those relevant to carcinogenesis and inflammation. The identification of CSC-affected genes, as well as the biological phenomena in which these genes participate, will**

allow generation of an atlas of specific molecular events caused by exposure to cigarette constituents. Eventually, these types of studies may be valuable in developing biomarkers of tobacco exposure and disease status in current and former smokers. In addition, such biomarkers may be useful in discriminating differential biological effects resulting from specific modifications to tobacco products. Finally, the finding that S9 affects the expression of a number of genes may have implications for a range of *in vitro* toxicogenetic assays that are used by regulatory agencies to evaluate potential harmful effects in exposed humans.